第一章　随机事件与概率

§1 样本空间与随机事件

1　随机现象

自然界中有许多现象在一定条件下必然会发生。例如：同性电荷必然互相排斥，在标准大气压下水加热到100℃必然沸腾等等，这类现象称为确定性现象，在一定条件下，必然出现的结果称必然事件。必然事件的对立面是不可能事件。

然而自然界中还存在大量的非确定性现象。例如观察某一商店每天来的顾客数与销售商品的数额都不是确定的；又如：正在放射α粒子的放射性物质，每天在同一规定的时间内放射的粒子数，事先无法确定，这类现象的共同点是：在基本条件保持不变的情况之下，可能出现这样的结果，时而又出现那样的结果，而且事先无法断言出现的究竟是哪一种结果，这类现象就称为随机现象。

2　随机试验

所谓随机试验。直观的讲，观察(或量测)在一定条件下随机现象出现的结果，即随机试验(简称试验)。进行一次试验就是在特定条件实现一次并观察其结果。在一次试验中，某个结果是否出现具有一定的偶然性。比如说，我们掷一次骰子，就可以看成是一次试验；因为掷骰子出现的点数是无法预先确定的，即试验的结果是偶然的、随机的。但许多实践早已证明：当进行大量的重复试验时，其结果就会出现某种固有规律性。

例如，在投掷一枚质地均匀的硬币时，只投掷一次时，投掷的结果是正面还是反面是无法确定的，但当大量重复投掷硬币，就可以看到出现正面的次数约占总试验次数的一半。又如某人打靶射击，若射击次数不多，靶上的弹着点似乎是随意分布的，但倘若进行大量的重复射击时，弹着点的分布就逐渐呈现规律性：它们大体上关于靶中心对称，靠近靶心的弹着点密，偏离靶心越远弹着点越稀少，且弹着点落在靶任意指定区域内的次数与射击次数 n 之比(频率)大体上保持稳定，且 n 越大，其频率稳定性就愈加明显，这种在大量重复试验中随机现象所呈现的固有规律，我们通常称之为统计规律。

为了研究的方便，我们有时也会把具有固定结果的试验，看成是随机试验的极端情形。有时，又需要把几次试验作为一个整体合起来看成一次随机试验，例如：可以把连续掷三次骰子看成是一次随机试验。

若试验具有下列共同特征：

1)　在相同的条件下，试验可重复进行；

2) 试验的一切可能结果是预先可以明确的，但每次试验前无法预先断言究竟会出现哪个结果。

则称之为随机试验，简称试验，记作 E 或 E_1，E_2 等。

例如：

E_1：掷一枚硬币，观察正面或反面出现的情况。

E_2：记录某一个服务台在 8∶00～9∶00 之间到来的顾客数。

E_3：在有噪声干扰的条件下，测量线路某一终端电压。

3 样本空间

对于随机试验 E，以ω表它的一个可能出现的试验结果，称ω为 E 的**一个样本点**。样本点的全体称为**样本空间**，用 Ω表示。即 $\Omega=\{\omega\}$。

从集合论的观点看，样本空间 Ω是由一切可能结果所构成的集合，而每个样本点 ω 是集合 Ω中的元素。对于上面的随机试验有：

E_1：掷一枚硬币一次，观察出现正、反面的情况。则 E_1 的样本空间为 $\Omega=\{\omega_1,\omega_2\}$；其中 ω_1 表示出现的是正面， ω_2 表示出现的是反面， ω_1 与 ω_2 分别为 Ω的样本点。

E_2：服务台在 8∶00～9∶00 之间到来的顾客数，则 $\Omega=\{n \mid n=0,1,2\cdots\}$；可见 Ω包含可列无穷多个样本点。

E_3：在有噪声干扰下，测量某终端电压，则 $\Omega=\{x \mid x\in R\}$，其样本点有不可数无穷多个。

注：在同样的试验条件下，由于试验的考察侧面与目的不同，可能选择不同的样本空间，这是初学者必须注意的。例如：E_4：一枚硬币投掷两次观察出现正面的次数(注意此时投掷两次硬币才算完成一次实验)。 $\Omega_4=\{0,1,2\}$，其 0,1,2 分别表示出现正面的次数，共有 3 个样本点。E_5：一枚硬币投掷两次观察出现正、反面的次序，则 $\Omega_5=\{\omega\omega, \omega\omega', \omega'\omega, \omega'\omega'\}$（其中：$\omega$代表出现正面 ，$\omega'$代表出现反面)，包含 4 个样本点。

为了便于理解，以下仅限制样本空间 Ω为可列集时，给出相应的概念。

4 随机事件

为了便于直观理解，不妨在这一小节先设样本空间 $\Omega=\{\omega\}$为可列集。

通常，对于某个随机试验来说，在一次试验中可能出现也可能不出现的事件，就称为**随机事件**。我们用大写英文字母 A、B、C、A_i 等来表示。例如：

在实验 E_1 中，$A_1=\{\omega_1\}$表示出现正面这一事件，$A_2=\{\omega_2\}$表示出现反面这一事件。A_1，A_2 都是事件。

在 E_2 中，$A=\{n \mid 0\leqslant n\leqslant 10\}$表示在 8∶00～9∶00 之间出现的顾客数不超过 10 人这一事件。

在 E_5 中，$A=\{\omega\omega, \omega'\omega)$表示掷第二次出现的是正面这事件。

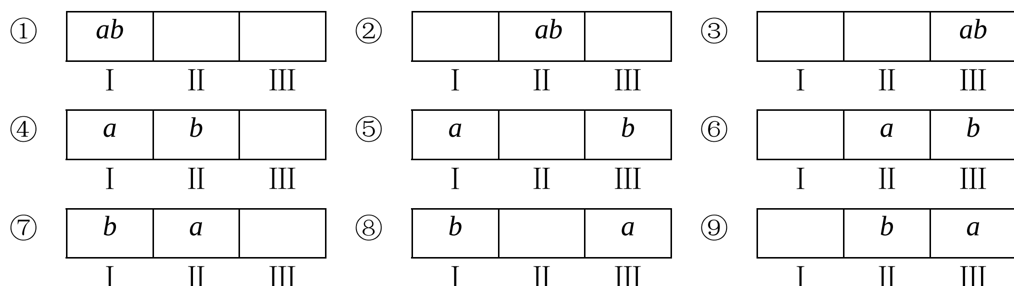
在引入了样本空间的定义后，即从集合论的观点看：粗略地说，样本空间 Ω的子集就是随机事件。

对于事件 $A\subset\Omega$，若 Ω 中的某一样本点 $\omega(\in A)$在本次实验中出现了，则称该次试验中事件 A 发生。若 $\omega\notin A$，即 ω 在本次实验中没有出现，则称该次试验中 A 不发生。

从随机事件的定义可以看出随机事件包含着两个极端情形。其一是在任何一次试验中必然出现的事件，称为**必然事件**。由于样本空间 Ω 在每次试验中必定出现(无论哪一个 ω 出现，均有 $\omega\in\Omega$) 故 Ω 是**必然事件**。另一种是在固定的试验条件下不能发生的事件，称为**不可能事件**，用空集∅表示不可能事件。

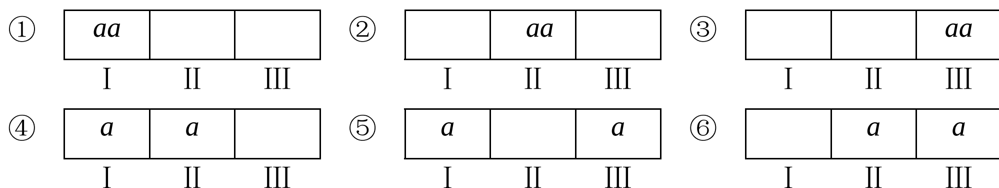
例 1.1 将一个红球 a，一个白球 b，随机地放在编号分别为Ⅰ、Ⅱ、Ⅲ的盒子中，观察两个球放置的所有可能结果，如下图用竖线表示盒壁，

把全部结果表示如下：



若用记号 ω_k $(1\leqslant k\leqslant 9)$表示上面第 k 个结果（样本点）。记：A=至少有一个盒放入两球，B=第一盒有球，C=第二盒无球，D=每盒恰有一球，E=至少有一盒是空的；则 $A=\{\omega_1, \omega_2, \omega_3\}$，$B=\{\omega_1, \omega_4, \omega_5, \omega_7, \omega_8\}$，$C=\{\omega_1, \omega_3, \omega_5, \omega_8\}$，$D=\varnothing$，$E=\Omega$。

例 1.2 若将二个不可辩的球(例如两个同样的红球)，随机地放在偏号分别为Ⅰ、Ⅱ、*III* 的盒子中，观察其放置的可能结果，若球用 a 表示，则其样本点为：



若用 $\omega_k(1\leqslant k\leqslant 6)$表示上面第 k 个样本点，记：A=至少有一盒有 2 球，B=第一盒有球，C=第二个盒是空的，则相应的事件可以表示为：

$A=\{\omega_1,\ \omega_2,\ \omega_3\}$，$B=\{\omega_1,\ \omega_4,\ \omega_5\}$，$C=\{\omega_1,\ \omega_3,\ \omega_5\}$。

再如：上一例中球与盒都是不可区分的，则上例中 $\omega_1\sim\omega_3$，三个样本点不再区别，同样第 4～第 6 个(即 $\omega_4\sim\omega_6$)样本点也不再区别。于是上例中的 6 个样本点，在本例中就分为二类即“两个球放在同一盒中”和“两个球放在不同盒中”，即本例应当有两个样本点 ω=两球放在同一盒中，ω'=两球放在不同盒中。

对于给定的一个随机试验，如何确切而恰当的建立与表示其样本空间，有时是建模和解决问题的重要一环，这一点在讨论古典概型时务请读者注意。

5 事件的关系与运算

从集合论的观点看，事件既然是一些集合，就必然存在着事件之间的关系，以及由若干事件来定义(或表示)的新的事件——即事件的一些运算及其规则。在以下的叙述中，设 Ω是给定的样本空间，A、B、C、A_i、B_i、……均表示为其中的事件。

（1）包含关系：

若事件A发生必然导致事件B发生，则称事件B包含事件A，记作$A\subset B$(或$B\supset A$)。由于事件A、B均是Ω的子集，因此从集合论的观点看来，若$\forall\ \omega\in A$有$\omega\in B$，则$A\subset B$。

（2）等价关系：

如果$A\subset B$且$B\subset A$，则称A与B等价，或称A等于B，记作$A=B$。

即$\forall\ \omega\in A$，必有$\omega\in B$，而且$\forall\ \omega\in B$，必有$\omega\in A$；$A=B$表示A与B是同一事件。在概率论中，对同一事件给出不同的等价表示形式是一种重要的技巧。

（3）事件的交：

定义$A\cap B$ 为事件A,B同时发生，称$A\cap B$为A与B的交，简记作AB；即 $A\cap B=\{\omega|\ \omega\in A$ 且 $\omega\in B\}$。

推广到 n 个事件：定义$\bigcap_{k=1}^{n}A_k$ 为事件 $A_1,A_2,\cdots,A_n$ 同时发生，称$\bigcap_{k=1}^{n}A_k$ 为事件$A_1,A_2,\cdots,A_n$的交，即$\bigcap_{k=1}^{n}A_k=\{\omega|\ \omega\in A_k,1\le k\le n\}$。记$\bigcap_{k=1}^{\infty}A_k$ 为事件$A_1,\cdots,A_k,\cdots$同时发生，称$\bigcap_{k=1}^{\infty}A_k$ 为$A_1,\cdots,A_k,\cdots$的可列交，即 $\bigcap_{k=1}^{\infty}A_k=\{\omega|\ \omega\in A_k,k\ge 1\}$。

（4） 互不相容：

$A,B\subset\Omega$，如果A,B不可能同时出现，即$AB=\varnothing$，则称二事件互不相容(或称互斥、互不相交)。

（5）事件的并：

定义$A\cup B$ 为事件A,B中至少有一事件发生，称$A\cup B$为A与B的并。

即 $A\cup B=\{\omega|\ \omega\in A$ 或 $\omega\in B\}$。

推广到 n 个事件：记$\bigcup_{k=1}^{n}A_k$ 为事件 A_k （$k=1,\ldots,n$）至少有一个事件发生，则称$\bigcup_{k=1}^{n}A_k$ 为事件$A_1,A_2,\cdots,A_n$的并，因此$\bigcup_{k=1}^{n}A_k=\{\omega|\ \omega\in A_1$, 或 $\omega\in A_2,\cdots$,或 $\omega\in A_n\}$。

同样，定义$\bigcup_{k=1}^{\infty}A_k$ 为“事件$A_1,A_2,\cdots,A_n,\cdots$ 至少有一个发生”，称$\bigcup_{k=1}^{\infty}A_k$ 为“$A_1,A_2,\cdots,A_n,\cdots$”的并。

若A、B互不相容时，记$A\cup B=A+B$，称为A与B的和。

推广：若$A_1,\cdots,A_k,\cdots$ 互不相容，则记 $\bigcup\limits_{k=1}^{n}A_k\triangleq\sum\limits_{k=1}^{n}A_k$， $\bigcup\limits_{k=1}^{\infty}A_k\triangleq\sum\limits_{k=1}^{\infty}A_k$。

（6）对立事件：

定义 $\overline{A}$ 为 A 不发生的事件，称事件 $\overline{A}$ 为事件 A 的逆, 或称 $\overline{A}$ 为 A 的对立事件。

有 $\overline{A}=\{\omega|\omega\in\Omega$，但 $\omega\notin A\}$。若记 $B=\overline{A}$，则 $B=\overline{A}\Leftrightarrow\begin{cases}A\cup B=\Omega\\ AB=\varnothing\end{cases}$。

（7） 事件的差：

“事件 A 发生，但事件 B 不发生”也是一个事件，记为 $A\backslash B$，称 $A\backslash B$ 为事件 A 与 B 的差。即 $A\backslash B=\{\omega|\omega\in A$ 且 $\omega\notin B\}$。 若 $B\subset A$，记 $A\backslash B\triangleq A-B$。

事件的运算规则：

1° 交换律：$A\cap B=B\cap A$， $AB=BA$；

2° 结合律：$A\cap(B\cap C)=(A\cap B)\cap C$， $A(BC)=(AB)C$；

3° 分配律： $A(B\cap C)=AB\cap AC$，$(A\cap B)C=(AC)\cap(BC)$，$A(B\backslash C)=(AB)\backslash(BC)$；

4° 德・摩根(*De-Morgan*)定律：$\overline{A\cup B}=\overline{A}\cap\overline{B}$， $\overline{A\cap B}=\overline{A}\cup\overline{B}$。

以上公式可以推广到有限个事件及可列个事件：

$$A(\bigcup_{k=1}^{\infty}B_k)=\bigcup_{k=1}^{\infty}(AB_k)，A\cup(\bigcap_{k=1}^{\infty}B_k)=\bigcap_{k=1}^{\infty}(A\cup B_k)，\overline{(\bigcup_{k=1}^{\infty}A_k)}=\bigcap_{k=1}^{\infty}\overline{A_k}，\overline{(\bigcap_{k=1}^{\infty}A_k)}=\bigcup_{k=1}^{\infty}\overline{A_k}。$$

我们只证明分配律和德・摩根定律，其余留给读者做为练习。

分配律：$A(B\cup C)=AB\cup AC$

证明： $\left.\begin{array}{l}\text{对}\ \omega\in A(B\cup C)\Leftrightarrow\omega\in A\ \text{且}\ \omega\in(B\cup C)\\ \text{而}\ \omega\in(B\cup C)\Leftrightarrow\omega\in B\ \ \text{或}\ \omega\in C\end{array}\right\}$

$$\Rightarrow\omega\in A(B\cup C)\Leftrightarrow\omega\in A\ \text{且}\ \omega\in B\ \text{或}\ \omega\in C$$

即 $\omega\in AB$ 或 $\omega\in AC\Leftrightarrow\omega\in AB\cup AC$，故：$A(B\cup C)=AB\cup AC$

德・摩根定律：$\overline{A\cup B}=\overline{A}\cap\overline{B}$。

证明： 对于 $\omega\in\overline{A\cup B}\Leftrightarrow\omega\notin A\cup B\Leftrightarrow\omega\notin A$ 且 $\omega\notin B\Leftrightarrow\omega\in\overline{A}$ 且 $\omega\in\overline{B}\Leftrightarrow\omega\in\overline{A}\cap\overline{B}$

故 $\overline{A\cup B}=\overline{A}\cap\overline{B}$。

6 事件序列

设 $A_k\subset\Omega$， $k=1,2,\cdots\cdots$，称 $\{A_k,k\geq1\}$ 为事件序列。若 $\forall n=1,2,\cdots$，$A_n\subset A_{n+1}$，则称之为单调增序列，定义：$\underset{n\to\infty}{Lim}A_n\triangleq\bigcup_{n=1}^{\infty}A_n$，称 $\underset{n\to\infty}{Lim}A_n$ 为 $\{A_n, n\geq1\}$ 的极限。若 $\forall n=1,2,\cdots$， $A_n\supset A_{n+1}$，则称之为单调减序列，定义 $\underset{n\to\infty}{Lim}A_n\triangleq\bigcap_{n=1}^{\infty}A_n$，称 $\underset{n\to\infty}{Lim}A_n$ 为 $\{A_n,n\geq1\}$ 的极限。 单调增（事件）序列和单调减（事件）序列统称为单调事件序列。

对于一般单调事件序列$\{A_n, n\geq 1\}$，令 $B_n=\bigcup_{k=n}^{\infty} A_k$，$C_n=\bigcap_{k=n}^{\infty} A_k$，显然 $B_n\supset B_{n+1}$，$C_n\subset C_{n+1}$，因此 $\{B_n, n\geq 1\}$和$\{C_n, n\geq 1\}$有极限。

令 $\overline{\lim_{n\to\infty}} A_n \triangleq \lim_n \sup A_n = \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n$ 称之为$\{A_n, n\geq 1\}$的上极限，它由属于无穷多个 A_n 的样本点构成。

令 $\underline{\lim_{n\to\infty}} A_n \triangleq \lim_n \inf A_n = \bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} A_n$，称之为$\{A_n, n\geq 1\}$的下极限，它由自某个 m 后属于所有 $A_n(n\geq m)$的样本点构成，易证恒有 $\underline{\lim_{n\to\infty}} A_n \subset \overline{\lim_{n\to\infty}} A_n$。

当 $\overline{\lim_{n\to\infty}} A_n = \underline{\lim_{n\to\infty}} A_n$ 时，称 $\lim_{n\to\infty} A_n = \overline{\lim_{n\to\infty}} A_n$ 为$\{A_n, n\geq 1\}$的极限。

例 1.3 设事件 $A, B\subset\Omega$，令 $A_{2n+1}=A$，$A_{2n}=B, n\geq 1$。

则：$\overline{\lim_{n\to\infty}} A_n = \bigcap_{k=1}^{\infty}\bigcup_{n=k}^{\infty} A_n = A\cup B$，$\underline{\lim_{n\to\infty}} A_n = \bigcup_{k=1}^{\infty}\bigcap_{n=k}^{\infty} A_n = AB$。

7 事件的示性函数

对于 $A\subset\Omega$，考虑定义在样本空间 Ω上的实函数：

$$I_A(\omega)=\begin{cases}1 & \text{当}\ \omega\in A\ \text{即}A\text{发生}\\ 0 & \text{当}\ \omega\notin A\ \text{即}\overline{A}\text{发生}\end{cases} \tag{1.1}$$

称 $I_A\triangleq I_A(\omega)$为事件 A 的示性函数。

引入事件的示性函数，可使事件的表示、事件的关系与运算，转化为数量的表示、数量的关系与运算。它是最简单的概念与记号，但它却扮演重要角色，尤其在现代概率论中发挥巨大的基础作用。容易验证：

$$\{I_A=1\}\triangleq\{\omega|\ I_A(\omega)=1\}=A,\ \{I_A=0\}\triangleq\{\omega|\ I_A(\omega)=0\}=\overline{A};$$

并且容易得到以下关系：$\forall\omega\in\Omega$ 有

(1) $A\subset B\Leftrightarrow I_A(\omega)\leq I_B(\omega)$；$A=B\Leftrightarrow I_A(\omega)=I_B(\omega)$；

(2) $I_{A\cap B}(\omega)=I_A(\omega)+I_B(\omega)-I_{AB}(\omega)$；$I_{A+B}(\omega)=I_A(\omega)+I_B(\omega)$；

(3) $I_{AB}(\omega)=I_A(\omega)\,I_B(\omega)$；

(4) $I_A(\omega)+I_{\overline{A}}(\omega)=1$；

(5) $I_{A-B}(\omega)=I_A(\omega)-I_B(\omega)$；

(6) 若：记 $a\vee b=max(a, b)$，$a\wedge b=min(a,b)$，$\bigvee_{k=1}^{n} a_k=\max_{1\leq k\leq n}\{a_k\}$，$\bigwedge_{k=1}^{n} a_k=\min_{1\leq k\leq n}\{a_k\}$，

则 $I_{A\cap B}(\omega)=I_A(\omega)\vee I_B(\omega)$，$I_{AB}(\omega)=I_A(\omega)\wedge I_B(\omega)$。$I_{\bigcup_{k=1}^{n}A_k}(\omega)=\bigvee_{k=1}^{n}I_{A_k}(\omega)$，

$I_{\sum_{K=1}^{n}A_k}(\omega)=\sum_{k=1}^{n}I_{A_k}(\omega)$，$I_{\bigcap_{k=1}^{n}A_k}(\omega)=\bigwedge_{k=1}^{n}I_{A_k}(\omega)$。

例：将 n 只球(1～n 号)随机地放进 n 只盒子(1～n 号)中去，一只盒子装一只球，若一只球装入与球同号的盒子中，称为一个配对，记 X 为总的配对数，X 如何表示呢?

我们自然而然地会想到示性函数，既然有配对与不配对的区分，则可以设：

$$X_i=\begin{cases}1 & \text{第 i 个球和盒子配对}\\ 0 & \text{第 i 个球和盒子不配对}\end{cases}$$

可见 X_i 为一个示性函数，而所求 X 可以表示成为 $X=\sum_{i=1}^{n}X_i$。

§2 杆 概率的公理化定义与性质

对于随机试验，人们不仅仅关心实验出现的可能结果，往往更关心各种结果出现的可能性究竟有多大。例如：大批量产品，随机抽取 3 个检验，发现次品的可能性有多大？一部设备运输时出现故障的可能性有多大？甲、乙两队正在进行激烈角逐，甲队问鼎的可能性有多大?

既然事件在每次试验中出现的可能性有大有小，因此就可用一个数字来表示事件发生的可能性，而这个数字就称为该事件的概率。这就是事件概率的直观意义。然而，对概率的确切描述却经过了一个漫长的历程，直至 1933 年才由前苏联著名数学家柯尔莫哥洛夫(*Kolmogorov*)给出概率的公理化定义；从而使概率论这门学科有了一个严格的理论基础，并在现代科学技术的推动和刺激下得以迅猛的发展。

为了能迅速而正确的理解和掌握事件的概率这一基本概念。我们先从概率的直观背景入手，而后再对一些特殊的随机模型讨论，最后给出一般情形的公理化定义。

1 古典概型

定义 2.1 若随机试验 E 及其样本空间Ω有下列二特性：

（1）样本空间Ω={ $\omega_1, \omega_2,..., \omega_n$}只有有限个样本点
（2）试验中每个样本点出现的可能性相同(等可能性)。
则称定义在该样本空间Ω上的概率模型为**古典概型**。

这类随机现象由于具有直观，简单的特点，所以在概率论发展的初期是人们研究的主要对象，先讨论它有助于理解概率论的许多基本概念，同时古典概型本身也仍然有着重要的应用背景。

定义 2.2 设随机试验 E 为古典概型，其样本空间为Ω={ $\omega_1, \omega_2,.., \omega_n$}；对于任一事件 A，其概率定义为：

$$\mathrm{P(A)} = \frac{\text{A中包含的样本点数}}{\Omega\text{中包含的样本点数}}。\qquad (2.1)$$

若记μ(A)={事件 A 中包含的样本点数}，则 $P(A)= \mu(A)/\mu(\Omega)$。

由定义，$P(A)$有以下性质：
（1） 非负性：$\forall A\subset\Omega$，$P(A)\geq 0$；
（2） 规一性：$P(\Omega)=1$；
（3） 设事件 $A_1,\cdots,A_m$ 互不相容，则：$P(\sum_{i=1}^{m}\mathrm{A}_i) = \sum_{i=1}^{m}\mathrm{P}(\mathrm{A}_i)$；
（4） 对于每一样本点 $\omega_i\,(1\leqslant i\leqslant n)$，有

$$P(\{\omega_1\}) = P(\{\omega_2\}) = \cdots\cdots = P(\{\omega_n\}) = \frac{1}{n}。$$

证明：只证明 (4)，其余的留作练习。

设事件 A_i={ ω_i}，$1\leqslant i\leqslant n$，则由定义知：$P(A_i) = \frac{1}{n}$，所以 $P(\{\omega_1\}) = P(\{\omega_2\}) = \cdots\cdots$ $= P(\{\omega_n\}) = \frac{1}{n}$。事实上，“每一个样本点出现等可能”与(2.1)式等价，即事件 A 的概率 $P(A)$仅与 A 中所含的样本点数有关，而与它具体含有的是哪些样本点无关。

例 2.1 E_1：掷一枚硬币一次，观察出现正、反面的情况。只有两个样本点，即 n=2；又由于硬币的均匀对称性，故二个样本点 ω（出现正面）、ω'（出现反面）是等可能的，故 E_1 是古典概型。则 $P(\{\omega\})=P(\{\omega'\})=1/n=1/2$。

例 2.2 E_2：掷一颗骰子观察出现的点数。有 6 个样本点（n=6），假定骰子是均匀对称的，则可认为每个样本点 $\omega_i\,(1\leqslant i\leqslant 6)$出现的可能性相同，从而 E_2 为古典概型。
则 $P(\{\omega_i\})=1/n=1/6\ (1\leqslant i\leqslant 6)$。若事件 A 表示出现偶数点，事件 B 表示出现点数至少为 3，则 A={ $\omega_2, \omega_4, \omega_6$} 即 k=3，所以 $P(A)= k/n=3/6=1/2$；B={ $\omega_3, \omega_4, \omega_5, \omega_6$}，即 k=4，所以 $P(B)= k/n=4/6=2/3$。

例 2.3 袋中有 10 球，6 个白球，4 个红球。假设每个球被取出是等可能的，现随机的取出 3 个，试求取出 k 个红球的概率 ($0\leq k\leq 3$)。

解：将球编号，记 6 个白球依次为 1~6，4 个红球为 7~10。假设每球被取出等可能，这等价于从 10 个号码中任取 3 个，不计次序的每一种取法是等可能的。因而若每次随机的取出 3 个，只观察其号码，不计其次序，则一个样本点等价于从 10 个号码中取其 3，

不计次序的一种组合。故样本点数 $\mu(\Omega)=C_{10}^3$。记 A_k ={取 3 个，取出 k 个红球}($0\le k\le 3$)。显然，A_k 发生，当且仅当 3 个球中，需从 7~10 号中取出 k ($0\le k\le 3$) 个，且从 1~6 号中取出 $(3-k)$ 个。由乘法原理有：$\mu(A_k)=C_4^k C_6^{3-k}$ ($0\le k\le 3$)，得 $P(A_k)=C_4^k C_6^{3-k}/C_{10}^3$。

注意：

（1）在解上题中，将球编号，目的是为了使我们建立的样本空间中的每一个样本点出现是等可能的。倘若对此题，球不加编号，那么在题设的条件下，随机取出 3 个，仅观察取到红球的个数。此时如果相应建立的（即解题者脑子中设想的）样本空间为 $\Omega=\{A_0,A_1,A_2,A_3\}$。显然上述样本空间的每个样本点 A_0,A_1,A_2,A_3 不是等可能的。这时，如硬要套用古典概型的定义来求解就会发生错误。

（2）对于“等可能”的确切含义，要依据具体的题设条件给予确切的理解与描述。这是能否正确解决古典概型问题的关键之一。务请初学者倍加注意。

例 2.4 n 个人抽签分配 n 张彩票，设 n 张彩票中有 m 张有奖彩票($1\le m<n$)。假设每人抽到每一张彩票是等可能的。试求第 k 个人 ($1\le k\le n$)抽到有奖彩票的概率。问第 1 个人抽到有奖彩票和第 n 个人抽到有奖彩票的概率是否相等。

解：记 A_k =第 k 个人抽到有奖彩票($1\le k\le n$)。为求 $P(A_k)$设想将彩票编号 1~n，试验只观察第 k 个人抽彩票的结果。建立样本空间如下：Ω中的一个样本点对应于从 1~n 张彩票中取出一张。显然，$\mu(\Omega)=n$，$\mu(A_k)=m$ 。故：$P(A_k)=m/n$。($1\le k\le n$)。由上知：$P(A_k)=m/n$ 与 k 无关。得 $P(A_1)=\cdots=P(A_n)=\dfrac{m}{n}$。

以上结果表明，抽到有奖彩票的概率与抽签次序无关，这正是人们直观上感到抽签分配是“公平的”的理论解释。

2 几何概型

对于属于古典概型的随机事件，我们利用等可能性在有限样本空间上定义了任一事件 A 的概率。但是由于有“有限个等可能样本点”这一条件的限制，在涉及有“无限多个可能的样本点”的随机实验时，古典概型就不再适用。然而具有无限多个等可能的样本点的随机实验在现实中是随处可见的。

比如我们要乘坐一班长途客车，客车站每半小时发一趟车，即 12∶00～12∶30 之间任何一刻到达的可能性都是相等的，就是说这个随机试验具有无穷多个等可能的结果，它已不再属于古典概型。对于每个单个的实验结果的出现，比如说，客车恰在 12∶15 到来的可能性只能说是 0。但客车在 12∶00～12∶15 到达的概率是多大呢?为了求这些问题的解，我们把 12∶00～12∶30 看作数轴上的一段区间，而 12∶00～12∶15 就成为

这个区间的一个子区间，在这种情况下等可能性就表现为：所取的点落在区间内的可能性仅与该区间长度有关，而与区间的形状与位置无关。简而言之，我们把古典概型的样本空间有有限多个样本点的情形推广到样本空间有无限多个样本点的情况，而每一样本点仍保持着某种等可能性，此时得到了另一种随机试验的数学模型。这就是所谓的几何概型，先看几个简单的例子。

例 2.5 在一维直线[0，10]区域上投掷一质点，质点随机地落入其中，且落在[0,10]上每一点的可能性相同，求质点落在[3，5]区域上的概率。

解：由题意，所有样本点充斥的区间为[0，10]，所以 Ω=[0，10]，令 A={点落在[3，5]}，因此事件 A 的概率 $P(A)$ 自然定义为“[3, 5]区间的长度”与“[0, 10]区间的长度”的比值。即

$$P(A)=\frac{\mu(A)}{\mu(\Omega)}=\frac{\text{A中包含的线段长度}}{\Omega\text{中包含的线段长度}} \tag{2.2}$$

则 $P(A)=(5-3)/(10-0)=1/5$

易知，事件 A 的概率仅与 A 的长度有关，而与 A 的形状和位置无关，这与质点随机地落在[0,10]上每一点等可能，两者是等价的。

但由实变函数论可知，上述 $P(A)$ 的定义(2.2)是不严格的，因为在直线上还存在着不能定义“长度”的点集，即存在着不可测的点集，因此式(2.2)只对长度可测的集合 A 才有意义。所谓可测集，直观地说就是可以定义“长度”的点集，这样我们就把所关心的事件严格地限制在 Ω 中所有可测的子集上。

例 2.6 在一个 6 万平方公里的海域里，有表面积约达 2000 平方公里的大陆架贮藏着石油。假设在这片海域里随机地选定一点钻探，问能找出有油的概率有多大?

解：假设在海域中的每点的钻探是等可能的，Ω=｛钻探的可能区域｝，事件 A=｛钻到有油的区域｝，则由等可能性应有定义：

$$\mathrm{P(A)}=\frac{\mu(\mathrm{A})}{\mu(\Omega)}=\frac{\text{A的面积}}{\Omega\text{的面积}} \tag{2.3}$$

所以 $P(A)$=2000/60,000=1/30。

此时，等可能性等价于 $P(A)$ 只与 A 的面积有关，而与 A 的形状和位置无关。

这里的概率 $P(A)$ 也只对 R^2 上能定义面积的点集（可测集）才有意义，因此我们所考虑的事件必须限制在二维可测集上才有意义。

以下对在一维点集上定义的长度，在二维点集上定义的面积，在三维点集上定义的体积，n 维点集上定义的超体积等通称为集合的测度。因为它们有共同点：是集合的函数，且取值非负，具有可加性。直观粗略的说，能够“度量”与定义测度的点集，称为可测集（在一定的度量“标尺”下）。

定义 2.3 设随机试验 E 的样本空间 $\Omega\subset R^n$ 是 R^n 中的可测子集，具有有限的测度 $\mu(\Omega)>0$（记号 $\mu(\Omega)$ 表示集合 Ω 的测度），记 Φ 为 Ω 中的可测子集全体。若 $\forall A\in\Phi$，定义：

$$P(A)=\frac{\mu(A)}{\mu(\Omega)} \tag{2.4}$$

规定 $\varnothing\in\Phi$，且 $\mu(\varnothing)=0$。称 A 为事件，$P(A)$ 为事件 A 的概率。称 (Ω, Φ, P) 为**几何概型**。

显然，由于 A 的测度 $\mu(A)$ 是集合的函数，且取值非负，具有可加性。故由(2.4)式定

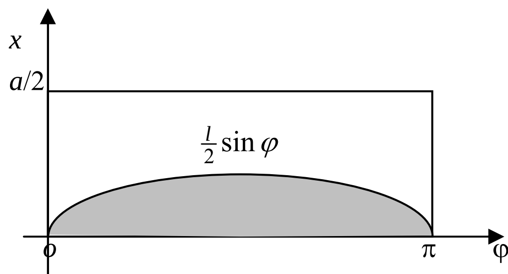
义的概率亦是集合的函数，且满足：(*i*) 非负性；(*ii*) 规一性；(*iii*) 可加性。同样，由(2.4)式定义的概率 $P(A)$蕴含着 Ω中各点的等可能性。

与前面例子类似，这里概率 $P(A)$只对可测集 A 才有定义，因此我们所考虑的事件必须限制在 Ω中的可测子集，而不是 Ω的所有子集，原因在于有不可测的子集存在。 几何概型在现实生活中有着典型而广泛地应用，下面再举一个例子。

例 2.6 (蒲丰(*Buffon*)问题) 平面中画着等距离的一些平行线。线与线之间的距离为 $a(a>0)$，向平面随机地投一长度为 $l\,(l<a)$的针，试求这针与任一直线相交的概率。

解：以 m 表示针的中点 x 与最近一平行线的距离，φ 表示针与平行线之间的夹角。易知：$\Omega=\{(x,\varphi)|0\le x\le \frac{a}{2},0\le\varphi\le\pi\}$，在 x，φ 平面上是一矩形。设事件 A={针与平行线相交}。则事件 A 发生的充要条件是：$x\le\frac{l}{2}\cdot\sin\varphi$

即 $A=\{(x,\varphi)\mid x\le\frac{l}{2}\cdot\sin\varphi,\ 0\le x\le a/2,\ 0\le\varphi\le\pi\}$，对应于下图中阴影面积。



这样我们的模型等价于向 Ω中等可能地投点的几何模型，求点落在区域 A 中的概率，因此由(2.4)式有：

$$\mathrm{P}(A)=\frac{\mu(A)}{\mu(\Omega)}=\frac{\int_0^{\pi}\frac{l}{2}\sin\varphi\,\mathrm{d}\varphi}{\frac{a}{2}\pi}=\frac{2l}{a\pi} \qquad (2.5)$$

3 频率与概率的统计背景

定义 2.5 设 E 为一随机试验，A 为一事件，在相同的条件下，把 E 独立的重复做 n 次，以$\mu_n(A)$表 A 在这 n 次试验中出现的次数，称 $f_n=\mu_n(A)/n$ 为事件 A 在这 n 次试验中出现的频率，而$\mu_n(A)$称为频数。

例2.7 历史上对于投掷一枚均匀硬币的实验有如下记录：

试验者	N	μ_n	f_n
Demorgan	2048	1061	0.5181
Baffon	4040	2048	0.5069

Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

其中 n 为实验次数，μ_n 代表出现正面的次数，$f_n=\mu_n/n$，从上述实验记录可以看出，随着 n 的逐渐增大，出现正面的频率 f_n 就越来越趋近于 1/2。

例 2.8 在对同一批产品进行质量抽验时，每次随机地抽取一件，其结果可能是合格品，也可能是次品，而且抽到合格品的可能性的大小是无法预知的。下面是某次抽验结果记录：

抽取件数 n	10	60	150	600	900	1200	1800
合格件数$\mu_n(A)$	7	53	131	548	820	1091	1631
合格频 $f_n(A)$	0.7	0.883	0.8730	0.913	0.911	0.909	0.906

随着每次抽取件数的增多，抽到是合格品的频率(即合格品件数与抽取产品个数之比)，总是越来越接近于某一确定值，即抽到合格品的频率呈现出一定的稳定性。上述记录表明，当抽取件数增多时，抽到合格品的频率稳定在 0.9 上下波动。

例 2.9 某国统计每年出生的男婴与女婴的资料表明，生男孩的频率大约在 0.517 左右徘徊，生女孩的频率约在 0.483 附近起伏。

以上试验说明：当 n 很大时，$f_n(A)$总是在某一实数 $P(A)$上下波动，而且随着 n 的增大，$f_n(A)$变化的总趋势是波动越来越小，且越来越趋近于 $P(A)$。许多实践也同样证明以上事实。而且，理论也早已证明：在相当广泛的条件下，当 $n\to\infty$ 时，频率 $f_n(A)$在一定意义下趋于常数 $P(A)$（详见第五章大数定律）。因此称 $P(A)$为事件 A 的概率。当 n 充分大时，可以用频率 $f_n(A)$作为 $P(A)$的近似值。频率的概念直观简单，容易掌握。因此可以根据频率的基本性质加以提炼与概括，作为定义事件概率的依据和直观背景。

从频率的定义可知事件 A 的频率 $f_n(A)$是集函数，具有下述性质：$0\leq f_n(A)\leq 1$，$f_n(\Omega)=1$。对于任意有限多个互不相容事件 A_1，A_2，…，A_m(即 $A_iA_j=\varnothing$，$\forall i\neq j$) 有 $f_n(\bigcup_{k=1}^{m}A_k)=\sum_{k=1}^{m}f_n(A_k)$，即具有可加性。

例 2.10 如果对例 2.6 中的投针试验重复进行 n 次，事件 A 发生(即投针与平行线相交)的次数为$\mu_n(A)$。当 n 根大时，其频率 $f_n(A)=\mu_n(A)/n$ 近似等于概率 $P(A)$，因此，若以 $f_n(A)$作为 $P(A)$的近似值代入（2.5）式，则得到 π 的近似值：

$$\pi\approx\frac{2\mathrm{nl}}{\mathrm{a}\mu_{\mathrm{n}}(\mathrm{A})}$$

用随机试验的方法确定圆周率 π，这真是奇妙的方法！

4 概率的公理化定义

在概率论发展早期，所研究的随机现象比较简单，大部分可以归入古典概型。对于这种模型，概率的计算可以通过某种可能性的假设进行。其结果也有相当明确的解释。这种成功使得人们试图用某种等可能性来定义概率，于是，由拉普拉斯给出的概率的古

典定义在整个十九世纪为人们所广泛接受。但是，这种定义的局限性很大，它既要求实验的可能结果总数有限，又要求具有某种等可能性，所以它的适用范围有限。对一般的随机现象明确的定义概率和其它基本概念，成了一个突出的问题。通过对事件与概率这两个基本概念的长期研究，人们发现事件的运算与集合的运算非常相似，概率与测度都是集函数，且有共同的基本性质：满足非负性与完全可加性。这一事实随着当时在实变函数论中关于勒贝格(*Lebesgue*)测度和积分的研究以及一般抽象测度和积分理论的发展而日益清晰起来。直到1933 年，前苏联数学家柯尔莫哥洛夫综合了前人结果，提出了概率论公理化结构，从而给出事件与概率的严格定义，推动了概率论的迅速发展。

下面就来叙述概率论中公理化结构。

（1） 集类

设 Ω 为一个抽象空间，即一个非空集合，它的元素称为点，一般用 ω 表示，由 Ω 中的某些点构成的集合用 A, B, A_1, A_2 等大写字母表示，它们都是 Ω 的子集。

$\omega\in A$ 表示 ω 为 A 的一个元素，称作 ω 属于 A。空集 ϕ 也是 Ω 的一个子集。由 Ω 中子集构成的集合称为集类，我们将用花写字母A, B, X等来表示。例A={ ϕ, Ω}，集类A中的元素是空集 ϕ 和全集 Ω。$A_1\subset\Omega, A_2\subset\Omega$，$A_1\neq A_2$, A={ A_1, A_2, ϕ}，集类A由三个元素构成。

我们用 2^{Ω} 表示由 Ω 的所有子集全体构成的集类。

（2） 概率空间三要素

（*a*） 样本空间Ω;

（*b*）σ-代数Φ：设Φ 是 Ω 的某些子集构成的非空集类，若满足以下条件：

(*i*) 若 $A\in\Phi$, 则 $\overline{A}\in\Phi$； (2.6)

(*ii*) 若对 $\forall k\geq 1, B_k\in\Phi$， 则 $\bigcup_{k=1}^{\infty}B_k\in\Phi$； (2.7)

则称Φ 为σ–代数。称（Ω，Φ）为可测空间。

（*c*）概率 P：P 是一个定义在Φ上的集函数，且它满足：

(*i*) $P(A)\geqslant 0$, 对$\forall A\in\Phi$，此即非负性；

(*ii*) $P(\Omega)=1$, 此即规一性；

(*iii*) 若 $A_i\in\Phi$，i=1,2,⋯，且 $A_iA_j=\varnothing(i\neq j)$，则

$$P(\sum_{i=1}^{\infty}A_i)=\sum_{i=1}^{\infty}P(A_i)\text{，} \tag{2.8}$$

此即可列可加性，也叫完全可加性。

则称 P 为可测空间（ Ω，Φ）上的一个**概率测度**（简称为概率）。称（ Ω, Φ，P）为概率空间，称Φ为事件域。若 $A\in$ ☞，称 A 为事件。称 $P(A)$为事件 A 的概率。

注：Ω中的一集合 A 是否是事件，取决于 A 是否属于Φ。在定义σ–域Φ时，一般并

没有要求Ω的全体子集都属于Φ，因而不是任何一集都一定是事件。但当Ω是有限或可列集时，如无特别指出，总是把Ω的全体子集 $2^{\Omega}=\{A, A\subset \Omega\}$作为Φ。

(3) σ–代数(也称σ–域)

下面举几个σ–代数的例子：

例 2.11 $\Phi_1=\{\phi, \Omega\}$，这是 Ω上最小的σ–代数。$\Phi_2=\{A|A\subset\Omega\}$是由 Ω的全体子集构成的，这是 Ω上最大的σ–代数。$\Phi_3=\{\phi, A, \overline{A}, \Omega\}$，其中 A 是 Ω的一个真子集。

命题 2.1 若Φ 为σ–代数，则：

(i) $\phi\in\Phi$, $\Omega\in\Phi$；

(ii) 若 $A_n\in\Phi, n\geq 1$, 则 $\bigcap\limits_{n=1}^{\infty}A_n\in\Phi$, $\bigcap\limits_{n=1}^{\infty}\bigcup\limits_{k=n}^{\infty}A_k\triangleq\overline{\lim\limits_{n\to\infty}}A_n\in\Phi$；

(iii) 若 $A, B\in\Phi$，则 $A-B\in\Phi$。

证明：留给读者作为练习。

从定义及性质表明, σ–代数对可列次并、交、差等所有运算是封闭的。这是σ–代数作为现代随机数学中基本概念和基本框架的主要原因之一。

可以证明，对于任意的一个集类X，一定存在包含X的最小σ–域。

定义 2.8 若Φ 是包含X的最小σ域，其中$X\subset 2^{\Omega}$，称Φ 是由X生成的σ域，记 $\Phi=\sigma(X)$。

例 2.12 $X_1=\{A, \Omega\}$, $X_2=\{A,B\}$, $\sigma(X_1)=\{\phi, A, \overline{A}, \Omega\}$, $\sigma(X_2)=\{A, \overline{A}, B, \overline{B}, \phi, \Omega, A\cap B, AB, \overline{A}\,\overline{B}, A\overline{B}, \overline{A}B, A\backslash B, \cdots\}$。

例 2.13 设: $A_1=\{[a,b), a,b\in R, a\leqslant b\}$; $A_2=\{(a,b], a,b\in R, a\leqslant b\}$; $A_3=\{[a,b], a,b\in R, a\leqslant b\}$; $A_4=\{(a,b), a,b\in R, a\leqslant b\}$; $A_5=\{(r_1, r_2), r_1, r_2$ 为有理数$\}$；$A_6=\{G: G$ 为 R 中开集$\}$； 以上集类生成相同的σ–代数，即$\sigma(A_1)=\sigma(A_k)$ $(1\leq k\leq 6)$（证明留作练习）。记$B=B^1=\sigma(A_1)$，称B为一维波雷尔（*Borel*）σ一域。B中的集称为**一维波雷尔（可测）集**

直观的说，$B=\sigma(A_1)$中包含一切开区间, 闭区间, 半开半闭区间，半闭半开区间，单个实数, 以及由它们经可列次交并差运算而得出的集合。

注：若样本空间Ω为有限集或可列集，如无特别声明，通常取Φ为Ω所有子集构成的集类。

(4) 概率的性质

从概率的公理化定义，我们可以推出概率的性质：

(i) $P(\varnothing)=0$；

证明： 显然有$\varnothing=\varnothing\cap\varnothing\cap\cdots$， $P(\varnothing)=\sum\limits_{i=1}^{\infty}P(\varnothing)$, 由概率非负性即得。

(ii) $P(\overline{A})=1-P(A)$；

请读者自己证明。

(*iii*) 有限可加性。 即若 $A_1, A_2, \cdots A_n \in \Phi$，且 $A_iA_j=\varnothing(i\neq j)$，则

$$P(\bigcup_{k=1}^{n} A_k)=\sum_{k=1}^{n} P(A_k)\text{。}$$

证明：由 $P(\varnothing)=0$ 及完全可加性可得。

(*iv*) $\forall A,B\in\Phi$，则：

$$P(A\backslash B)=P(A)-P(AB) \tag{2.11}$$

若 $B\subset A$，则：

$$P(A-B)=P(A)-P(B) \tag{2.11a}$$

证明：由 $A=AB+(A\backslash B)$，有：$P(A)=P(AB)+P(A\backslash B)$，得(2.11)。当 $B\subset A$ 时，有 $B=AB$，由(2.11)得(2.11a)。

(*v*) $P(A\cap B)=P(A)+P(B)-P(AB)$； (2.12)

证明：由 $A\cap B=A+(B\backslash A)$，有：$P(A\cap B)=P(A)+P(B\backslash A)$及(2.11)得(2.12)。

(*vi*) 若 $A\subset B$, 则 $P(A)\leqslant P(B)$；

证明：由 $A\subset B$，$B=A+(B-A)$，得 $P(B)=P(A)+P(B-A)\geq P(A)$。

(*vii*) $\forall A_k\in\Phi$，$1\leq k\leq n$，$n\geq 2$，

$$\begin{aligned}P(\bigcup_{k=1}^{n} A_k) &= \sum_{k=1}^{n} P(A_k) - \sum_{1\leq i<j\leq n} P(A_iA_j) + \sum_{1\leq i<j<k\leq n} P(A_iA_jA_k) + \cdots \\ &+\cdots+(-1)^{n-1}P(A_1A_2\cdots A_n)\end{aligned} \tag{2.13}$$

证明：由(2.12)及数学归纳法可得(2.13)。

推论：

$$P(\bigcup_{k=1}^{n} A_k)\leq \sum_{k=1}^{n} P(A_k) \quad\text{（有限次可加性）} \tag{2.14}$$

$$P(\bigcup_{k=1}^{n} A_k)\geq \sum_{k=1}^{n} P(A_k) - \sum_{1\leq i<j\leq n} P(A_iA_j) \tag{2.15}$$

(*viii*) 可列次可加性：

$$P(\bigcup_{k=1}^{\infty} A_k)\leq \sum_{k=1}^{\infty} P(A_k) \tag{2.16}$$

(*ix*) （概率连续性）若$\{A_n, n\geqslant 1\}$为单调事件序列，则

$$P(\lim_{n\to\infty} A_n)=\lim_{n\to\infty} P(A_n) \tag{2.17}$$

证明：若$\{A_n, n\geqslant 1\}\in\Phi$，且 $A_1\subset A_2\subset\cdots$，我们来证明上述性质。

记 $B_1=A_1, B_2=A_2-A_1, \cdots, B_n=A_n-A_{n-1}$，显然有 $A_n=\bigcup_{k=1}^{n} A_k=\sum_{k=1}^{n} B_k$，$\lim_{n\to\infty} A_n=\bigcup_{n=1}^{\infty} A_n=\sum_{n=1}^{\infty} B_n$

且$\{B_n, n\geqslant 1\}$两两不相交，由完全可加性有：

$P(\lim\limits_{n\to\infty} A_n)=P(\sum\limits_{n=1}^{\infty} B_n)=\sum\limits_{n=1}^{\infty} P(B_n)=\lim\limits_{n\to\infty}\sum\limits_{k=1}^{n} P(B_k)=\lim\limits_{n\to\infty} P(A_n)$，即得(2.17)

对$\{A_n, n\geqslant 1\}\in\Phi$，且$A_1\supset A_2\supset\cdots\supset A_n\supset\cdots$，注意到$\overline{A}_1\subset\overline{A}_2\subset\cdots$，$\bigcup\limits_{1}^{\infty}\overline{A}_n=\overline{(\bigcap\limits_{1}^{\infty} A_n)}$可得(2.17)。详细证明留给读者作为练习。

例 2.14 设样本空间$\Omega=\{\omega_1, \omega_2, \cdots \omega_n\}$仅有有限个样本点，$\Phi=\{A|A\subset\Omega\}$，$P(A)=\dfrac{\mu(A)}{\mu(\Omega)}$，其中$\mu(A)$是$A$的测度（具有非负、可加性的集函数）。则称$(\Omega, \Phi, P)$是一个概率空间，称为有限概率空间。

下面的例 2.15 初学者可跳过不看。

例2.15 设$\Omega=R$，Φ=B，B为一维波雷尔（*Borel*）σ一域。在可测空间$(\Omega, \Phi)=(R, \text{B})$上定义$P(\cdot)$满足：$\forall A\in\Phi=\text{B}$，$P(A)=\int\limits_{t\in A}(2\pi)^{-1}e^{-t^2/2}dt$。由于$P(\Omega)=\int\limits_{t\in R}(2\pi)^{-1}e^{-t^2/2}dt=1$（详见第二章正态分布），则可证明$(\Omega, \Phi, P)=(R, \text{B}, P)$是一概率空间。通常称$P$为一维标准正态概率测度。

例 2.16 1. [0,1]上的 *Borel* 概率空间。设Ω=[0,1]，Φ=B[0,1]，即B[0,1]是B局限在[0,1]上的 *Borel* σ-域。称(Ω, Φ)=([0,1], B[0,1])为[0,1]上的 *Borel* 可测空间。设在可测空间([0,1], B[0,1])上定义一概率测度P，它满足：当$\forall A=[a,b]\in$B[0,1]时，$P(A)=b-a$，称(Ω, Φ, P)=([0,1], B[0,1], P)为[0,1]上的 *Borel* 概率空间，称P为[0,1]上的 *Borel* 概率测度。

2. 令B=[0,1]上有理点全体，$\overline{B}$=[0,1]上无理点全体。

(1) 试证：$B\in\Phi$，$\overline{B}\in\Phi$；

(2) 用公理化定义与性质，求证：$P(B)=0$，$P(\overline{B})=1$。

证明：$\forall a\in[0,1]$，单点集$\{a\}=\bigcap\limits_{n=1}^{\infty}[a, a+\dfrac{1}{n})\in$B[0,1]= Φ，而$B=\{0\}\bigcup\{\frac{m}{n}:1\leq m\leq n, n,m=\{1,2,\cdots\}\}$是可列单点集的并，故$B\in\Phi$。且$\overline{B}=[0,1]-B\in\Phi$。又$\forall a\in[0,1]$，$P(\{a\})=0$，由完全可加性知$P(B)=0$，而$\overline{B}=[0,1]-B$，故$P(\overline{B})=1-0=1$。证毕。

§3 古典概型的计算

上一节中，给出了古典概型的定义，并举了一些简单例子；这一节重点讨论有关古典概型的计算。

古典概型是具有以下两点性质的试验：

（1）试验的样本空间的元素只有有限个。

（2）试验中每个样本点发生的可能性相同。

事件 A 的概率定义为：

$$P(A)=\frac{A\text{中包含的样本点数}}{\Omega\text{中包含的样本点数}}=\frac{\mu(A)}{\mu(\Omega)}。$$

为计算样本点数，往往需要用到排列组合的有关知识，而这部分内容中学已经讲过，这里只介绍简单组合的推广，即 n 个元素分为 k 组的组合数。

设有 n 个不同元素分为 k 组 $(2\le k\le n)$，第一组合 m_1 个，第二组含 m_2 个, 第 k 组含 m_k 个, $\sum_{i=1}^{k} m_i = n$。为计算其组合数，可以先将 n 个分为 m_1，$n-m_1$ 两个组；再将 $n-m_1$，分为 m_2, $n-m_1-m_2$ 两个组；如此类推，即可得组合数为：

$$\frac{n!}{m_1!(n-m_1)!}\cdot\frac{(n-m_1)!}{m_2!(n-m_1-m_2)!}\cdots\frac{(n-m_1-\cdots-m_{k-2})!}{m_{k-1}!(n-m_1-\cdots-m_{k-1})!}=\frac{n!}{m_1!m_2!\cdots m_k!} \quad （*）$$

种不同的分组方法（不计其次序）。

不难看出：上式是多项式 $(x_1+x_2+\cdots+x_k)^n$ 的展开式中 $x_1^{m_1}x_2^{m_2}\cdots x_k^{m_k}$ 项的系数，因此我们也称（*）式为多项式系数。

古典概型中事件概率的计算步骤大致为：

(1) 由给定的随机试验，建立合适的样本空间，判断其等可能性是否满足；

(2) 求出 Ω 及 A 中的样本点数 $\mu(\Omega)$ 及 $\mu(A)$;

(3) 求概率 $P(A)$。

分类举例如下：

1 抽球问题

例 3.1 箱中有 10 个球，其中 6 个白球，4 个红球，任取 3 个；

求：(*a*)．恰有一红球的概率；(*b*)．至少有一红球的概率。

解：将球编号：白球 1~6，红球 7~10。

设随机试验是任取 3 球，只考察号码，而不记其取出次序。一个样本点对应于从 10 个号码元素中取出 3 个，不计其次序的一种组合取法，则$\mu(\Omega)=C_{10}^3$。令：A_i={恰取到 i 个红球}，B={至少取出一个红球}；A_i 发生当且仅当：从 1–6 中取出 3–i 个，从 7–10 中

取出 i 个；故$\mu(A_i)=C_6^{3-i}C_4^i$。所以 $P(A_i)=\dfrac{C_6^{3-i}C_4^i}{C_{10}^3}$；于是 $P(A_1)=\dfrac{C_6^2C_4^1}{C_{10}^3}=\frac{60}{120}=\frac{1}{2}$。

又 $B=\Omega-A_0$，于是 $P(B)=1-P(A_0)=1-\dfrac{C_6^3}{C_{10}^3}=1-\dfrac{1}{6}=\dfrac{5}{6}$。

例 3.2 杆箱中有 10 个球，6 白 4 红，随机地一个个取出不放回，求第 k 次($1\leqslant k\leqslant 10$)取到红球的概率。

解：下面给出两种解法，读者可以从中体会建立样本空间的方法。

法一：将球如例 3.1 编号。因为所求为第 k 次取出红球的概率，故试验时只观察球一个个取出时第 k 次的结果。建立样本空间 Ω如下：Ω中的一个样本对应于从 10 个球中取出 1 个，故$\mu(\Omega)=C_{10}^1$。令 A_k={第 k 次取出红球}，则 A_k 中的样本对应于从 7~10 号中取出 1 个，故$\mu(A_k)=C_4^1$。所以 $P(A_k)=\dfrac{\mu(A_k)}{\mu(\Omega)}=\dfrac{C_4^1}{C_{10}^1}=\dfrac{4}{10}$ $(1\leq k\leq 10)$。

因此：第 k 次取到红球的概率与 k 无关。

法二：球不偏号，只区分红，白颜色。试验时，只考察 4 个红球占据有编号的 10 个位置中的哪 4 个位置。于是样本空间 Ω中的一个样本点对应着 10 个元素中取出 4 个，不计其次序的一种取法，故 $\mu(\Omega)=C_{10}^4$。A_k={第 k 各位置上是红球}，A_k 发生对应当且仅当"有一红球占据第 k 位置，而其余 3 个红球占据剩下的 9 个位置中的 3 个"，故 $\mu(A_k)=C_9^3$.，$\therefore P(A_k)=\dfrac{C_9^3}{C_{10}^4}=\dfrac{4}{10}$。

从上面两种方法可以看出，在解决同一具体问题时，可建立不同的样本空间。关键在于满足古典模型的两个条件并且适用于问题的解决。

例 3.3 题设同例 3.2；求在第 3 次时，首次取到红球的概率。

解：同例 3.2 将球编号，只观察前三次抽球情况，据此建立样本空间 Ω。Ω的每一个样本对应于从 10 个元素中取 3 个考虑次序的一种排列，故 $\mu(\Omega)=A_{10}^3$，令 A=第三次首次取到红球，则 A 发生当且仅当前 2 次从 1-6 号中取球，第 3 次从 7-10 号中取球，

故 $P(A)=\dfrac{A_6^2A_4^1}{A_{10}^3}=\dfrac{1}{6}$。

另法。同例 3.2 法二，球不编号。只考察 4 个红球放置的位置，此时 $\mu(\Omega)=C_{10}^4$。而事件 A 发生，当且仅当，其中一红球占据第三个位置，且其它 3 个红球占据从 4~10 中的 3 个，则 $\mu(A)=C_7^3$。得 $P(A)=C_7^3/C_{10}^4=1/6$。

实际生活中常常会遇到各种各样的“抽球”问题。比如把“白球”与“红球”换成“合格品”与“次品”,“正面”与“反面”,“成功”与“失败”,“有信号”与“无信号”等等。许多问题都能形象化地用抽球问题加以描述，所以抽球问题具有典型意义。

例 3.4 箱中有 a 个白球，b 个红球，采用有放回和无放回两种抽样方式从中取出 n（$n\leqslant a+b$）个球，恰有 k（$0\leqslant k\leqslant n$）个红球的概率各是多少？

解：由于抽样方式不同，即随机试验不同，故须分别讨论。

(a) 有放回抽样。此时，抽取 n 个的一种取法对应于从$(a+b)$个元素中取 n 个的可重复排列的一种，故 $\mu(\Omega)=(a+b)^n$；令 A={抽出的 n 个中恰有 k 个红球}，则 $\mu(A)=C_n^k a^k b^{n-k}$，$(0\le k\le n)$。故：$P(A)=C_n^k(\frac{b}{a+b})^k(\frac{a}{a+b})^{n-k}$。

(b) 无放回抽样。抽取 n 个球，不考虑其次序；此时，Ω中每一样本点对应于从$(a+b)$个元素中取出 n 个的一种组合，故 $\mu(\Omega)=C_{a+b}^n$，而恰有 k 个红球需 $0\le k\le b$,有

$\mu(A)=C_b^k C_a^{n-k}$，$0\le k\le b$，$1\le n\le a+b$，

故 $P(A)=C_b^k C_a^{n-k}/C_{a+b}^n$，$0\le k\le b$，$1\le n\le a+b$。

2 分房问题

例 3.5 有 n 个人，每个人都以同样的概率被分配在 N ($n\leqslant N$)间房中的每一间中，求下列事件的概率：

A：某指定 n 间房中各有 1 人；

B：恰有 n 间房，其中各有 1 人；

C：某指定房间中恰有 $m(m\leqslant n)$人。

解：设随机试验为：把每个人随机分配到 N 个间房中任一间，据此建立样本空间 Ω；由于每一人分到 N 个房间中有 N 种分法，由乘法原理：$\mu(\Omega)=N^n$。

① 今固定某 n 个间房且每间一人，第一人可分配到其中任一间，有 n 种分法，第二人可分配到余下 n–1 间中任一间，有 n–1 种分法，…，故 $\mu(A)=n!$，所以 $P(A)=\frac{n!}{N^n}$。

② 如果这 n 个间房可自 N 间中任意选出，那么只有 C_N^n 种选法，故 $\mu(B)=n!\cdot C_N^n$

所以 $P(B)=\frac{n!\cdot C_N^n}{N^n}=\frac{N!}{N^n(N-n)!}$。

③ 事件 C 中的 m 个人可自 n 个人中任意选出，故有种 C_n^m 种选法，其余$(n–m)$个人可任意分配到剩下的 N–1 间房里，共有 $(N-1)^{n-m}$ 种分法，$\therefore\mu(C)=C_n^m(N-1)^{n-m}$。

所以 $P(C)=\frac{C_n^m(N-1)^{n-m}}{N^n}=C_n^m(\frac{1}{N})^m(\frac{N-1}{N})^{n-m}$。

例 3.6 将 r 个质点（不可分辨）随机放入 n 个盒子中($1\leqslant r\leqslant n$)(假设每一种方法等可能)，求下列事件的概率：① A=“某指定的 r 个盒中各有一质点”， ② B=“某指定的一个盒中恰有 k 个质点” ($1\leqslant k\leqslant r$)。

解：① 用竖线“|”表示盒壁，用“*”表示质点，将盒子和质点并列排成一排，如 r=3，n=4 的一种放法表示为：| ** | | * | |，表示各个盒子中的质点数依次为 2，0，1，0。把每个壁与每个点都看成一个位置，其中两端的竖线固定，其余的 $n-1$ 条竖线和 r 个“*”依次占据 $n+r-1$ 个位置。令 Ω={ r 个质点的放法全体}，于是 r 个质点的一种占位就相当于 $n+r-1$ 个位置中 r 个被“*”占据的一种。等价与从($n+r-1$)个数中不计次序的取出 r 个数的一种取法，故$\mu(\Omega)=C_{n+r-1}^r$。事件 A 只是 r 个质点的一种分布，故$\mu(A)=1$，所以

$$P(A)=\frac{\mu(A)}{\mu(\Omega)}=\frac{1}{C_{n+r-1}^r}=\frac{r!(n-1)!}{(n+r-1)!}。$$

② B=“某指定的一盒中恰有 k 个质点”。这样，B 中的一个样本点就对应于有 k 个质点放在指定盒中，而其余的 $r-k$ 个质点放在剩下的 $n-1$ 个盒中的一种放法。

故由①可知：$\mu(B)=C_{(n-1)+(r-k)-1}^{r-k}=C_{n+r-k-2}^{r-k}$，所以

$$P(B)=\frac{\mu(B)}{\mu(\Omega)}=\frac{C_{n+r-k-2}^{r-k}}{C_{n+r-1}^r}。$$

本题与例 3.5 同属“分房问题”。只是在例 3.5 中人是可辨的，即人的占据不仅依赖于每个间房的人数，而还取决于不同人；而本题中，由于质点不可辨，所以其分布只依赖于盒的质点个数，这样在计算样本点数时，所用的方法就不相同，请读者仔细体会。

3 随机取数问题

例 3.7杆 从 1,2,⋯,10 这十个数字中任取一个(假定每个数字都以相同的概率被取中)，取后还原。先后取出 7 个数字，试求下列各事件的概率：

① A_1：7 个数字全不相同；

② A_2：不含 1 和 10；

③ A_3：10 恰好出现两次；

④ A_4：至少出现两次 10。

解：每取一次，有 10 种不同的取法；还原后，再取仍有 10 种不同的取法；因此由乘法原理，Ω中共有 10^7 个不同的样本点。

① A_1=“7 个数字全不相同”，容易看出$\mu(A)=10\times9\times8\times7\times6\times5\times4=10!/3!$。

所以 $P(A_1)=\dfrac{\mu(A_1)}{\mu(\Omega)}=\dfrac{10\times9\times8\times7\times6\times5\times4}{10^7}\approx0.06048$。

② A_2="不含 1 与 10"。A_2 中的一个样本点对应于“从 2~9 这 8 个数中任取 1 个，重复 7 次的一种取法”，故 $\mu(A_2)=8^7$；所以 $P(A_2)=\frac{8^7}{10^7}$

③ A_3=“10 恰好出现两次”。出现的两次 10 可以是 7 次中的任意两次，而其他 5 次，每次只能取剩下 9 个数字中的一个，故 $\mu(A_3)=C_7^2\cdot9^5$，$P(A_3)=C_7^2\cdot9^5/10^7$。

一般地，10 恰好出现 k 次(k≤7)的概率： $P_k=\dfrac{C_7^k\cdot9^{7-k}}{10^7}$。

④ A_4=“至少出现两次 10”。设 B_k=10 恰好出现 k 次(k=0, 1, 2, 3, 4, 5, 6, 7)，则它们彼此互不相容，所以 $A_4=\sum_{k=2}^7 B_k$；由③及概率的有限可加性，有

$$P(A_4)=\sum_{k=2}^{7}P_k=\frac{\sum_{k=2}^7 C_7^k\cdot9^{7-k}}{10^7}。$$

又 $A_4=\Omega-B_0-B_1$，所以 $P(A_4)=1-\sum\limits_{k=0}^{1}P_k=1-\dfrac{\sum_{k=0}^1 C_7^k\cdot9^{7-k}}{10^7}$。

从以上几类事件概率的计算，可以总结出，在求解有关古典概型的事件概率时应做到：

1．善于对较为复杂的事件进行分解和转化，变为简单易求概率的计算；
2．针对具体问题，建立适当的样本空间；对于复杂的问题可以籍助于直观表示；
3．善于利用排列组合的相关知识，正确求解 $\mu(\Omega)$，$\mu(A)$。

§4 条件概率与全概率公式

在实际问题中，有时不仅要知道事件 A 发生的概率 $P(A)$；而且还要知道在事件 B 发生的条件下，事件 A 发生的条件概率，记这一条件概率为 $P(A|B)$。由于引入“在事件 B 发生的条件下”这一前提，$P(A|B)$与 $P(A)$通常是不一样的。

例 4.1 投掷骰子，观察出现的点数。$\Omega=\{1,2,3,4,5,6\}$；若 A={点数为偶数}，B={所

掷点数大于 3}；易知 $P(A)=3/6=1/2$， $P(B)=3/6=1/2$。当 B 发生时，所有可能的样本点数为 3，在此条件下，A 发生的样本点数为$\mu(AB)=2$，故而 $P(A|B)=2/3$，$P(A) \neq P(A|B)$。

上例中 $P(A)$，$P(B)$称为无条件概率，$P(A|B)$称为条件概率。由此例可见条件概率与无条件概率并不相同，需进一步研究。

1 条件概率的定义

定义 4.1 设(Ω,Φ,P)为概率空间，$A,B\in\Phi$，设 $P(B)>0$，定义：

$$P(A|B)\triangleq\frac{P(\mathrm{AB})}{P(\mathrm{B})} \tag{4.1}$$

称 $P(\mathrm{A}|\mathrm{B})$ 为事件 B 发生下，事件 A 发生的条件概率。

为了区别 $P(\mathrm{A}|\mathrm{B})$，称 $P(A)$为事件 A 发生的无条件概率。

在古典概型中，若事件 B 中含有 m 个不同的样本点，AB 中含有 k 个不同样本点，而Ω中样本点数为 n。则由定义 4.1 知：$\mathrm{P(A|B)}=\frac{\mathrm{P(AB)}}{\mathrm{P(B)}}=\frac{k/n}{m/n}=\frac{k}{m}.$

在几何概型中，以 $\mu(A_i)$ 表示 A_i 的面积，$\mu(\Omega)$ 为总面积。则

$$\mathrm{P(A_1|A_2)}=\frac{\mathrm{P(A_1A_2)}}{\mathrm{P(A_2)}}=\frac{\mu(\mathrm{A_1A_2})/\mu(\Omega)}{\mu(\mathrm{A_2})/\mu(\Omega)}=\frac{\mu(\mathrm{A_1A_2})}{\mu(\mathrm{A_2})}$$ 。

在事件 B 给定时，条件概率 $\mathrm{P}(A|B)$ 也是 A 的集函数，请读者自己根据定义验证它具有非负性、规一性和可列可加性。这样条件概率具备概率的所有性质，如果条件概率中取 $B=\Omega$，则：$\mathrm{P}(A|\Omega)=\frac{\mathrm{P}(A\Omega)}{\mathrm{P}(\Omega)}=\mathrm{P}(A)$，因此无条件概率只是条件概率的一种特殊情况。

2 条件概率的乘法公式和全概率公式

这里给出条件概率的三个基本公式：乘法公式，全概率公式和贝叶斯公式。这些公式在随机数学中起着很重要的作用，现在分别介绍。

（1）乘法公式

定理 4.2 设 $A,B\in\Phi$，若 $P(B)>0$，则

$$P(AB) = P(B)P(A|B) \tag{4.2}$$

称之为乘法公式。

证明： ∵ $P(B)>0$，由条件概率定义即证。

推论 1 设 $A_i\in\Phi$，$i=1,2,3$，若 $P(A_1A_2)>0$，则 $P(A_1A_2A_3)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)$

证明： ∵$P(A_1)\geq P(A_1A_2)>0$，故：

$$P(\underbrace{A_1A_2}A_3)=P(A_1A_2)P(A_3|A_1A_2)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)。$$

类似地有一般的情形

推论 2 设 $A_i\in\Phi$，$1\le i\le n$，若 $P(A_1A_2\cdots A_{n-1})>0$，则

$$P(A_1A_2A_3\cdots A_n)=P(A_1)P(A_2|A_1)P(A_3|A_1A_2)\cdots P(A_n|A_1A_2\cdots A_{n-1})。$$

例 4.2 袋中有 10 个球，6 白 4 红，一个一个的取出；令 A_k=第 k 次取出红球，其中 $1\le k\le 10$；求：$P(A_1A_2), P(A_2)$。

解：方法一：按照古典概型的计算方法：

$$P(A_1A_2)=\frac{C_4^2}{C_{10}^2}=\frac{2}{15}，\quad P(A_2)=\frac{C_4^1}{C_{10}^1}=\frac{2}{5}。$$

方法二：$P(A_1A_2)=P(A_1)P(A_2\mid A_1)=\frac{4}{10}\times\frac{3}{9}=\frac{2}{15}$。由 $A_2=\Omega A_2=A_1A_2+\overline{A}_1A_2$，故：

$$P(A_2)=P(A_1A_2+\overline{A}_1A_2)=P(A_1A_2)+P(\overline{A}_1A_2)=P(A_1)P(A_2\mid A_1)+P(\overline{A}_1)P(A_2\mid\overline{A}_1)$$
$$=\frac{2}{15}+\frac{6}{10}\times\frac{4}{9}=\frac{2}{5}$$

在方法二求 $P(A_2)$的过程中，对Ω作如下分解：$\Omega=A_1+\overline{A}_1$，且使 A_1A_2 与 $\overline{A}_1A_2$ 互不相容；这种分解方法体现了以下全概率公式的基本思想。

（2） 全概率公式

定义 4.2 若$\{B_i, i\ge 1\}$满足：$B_i\in\Phi$，$\forall i\ne j, B_iB_j=\phi$，$\bigcup_{i=1}^{\infty}B_i=\Omega$；则称$\{B_i, i\ge 1\}$为样本空间Ω的一个分解。

定理 4.3 若$\{B_i, 1\le i\le n\}$为样本空间Ω的一个分解，$A\in\Phi$，则

$$P(A)=\sum_{i=1}^{n}P(B_i)P(A\mid B_i)，\tag{4.3}$$

称之为全概率公式。

证明：$\because A=A\Omega=A(\bigcup_{i=1}^{n}B_i)=\sum_{i=1}^{n}(B_iA)$，所以

$$P(A)=P(\bigcup_{i=1}^{n}(B_iA)) \quad （事件的等价性）$$
$$=\sum_{i=1}^{n}P(B_iA) \quad （有限可加性）$$
$$=\sum_{i=1}^{n}P(B_i)P(A\mid B_i) \quad （乘法公式）。$$

将事件 A 分解成若干互不相容的事件 AB_i 的和，从而将复杂事件 A 的概率 $P(A)$转化

为事件 AB_i 的概率 $P(B_iA)$的和。

下面几个例子说明全概率公式在简化计算方面的作用。

例 4.3 从 0~9 中任取两数，求其和大于 10 的概率。

解： 记 A={0~9 中和大于 10 两数}，B_i={第一次取出数 i}，（$1\le i\le 9$），则 $\forall i\neq j, B_iB_j=\phi$，$\bigcup_{i=0}^{9}B_i=\Omega$。又 $P(B_i)$=1/10；$P(A|B_0)$=0；$P(A|B_1)$=0；当 $2\le i\le 5$ 时, $P(A|B_i)=(i-1)/9$；当 $6\le i\le 9$ 时,$P(A|B_i)=(i-2)/9$；所以

$$P(\mathrm{A})=\sum_{\mathrm{i}=0}^{9}P(\mathrm{B_i})P(\mathrm{A}|\mathrm{B_i})=\sum_{\mathrm{i}=2}^{5}P(\mathrm{B_i})P(\mathrm{A}|\mathrm{B_i})+\sum_{\mathrm{i}=6}^{9}P(\mathrm{B_i})P(\mathrm{A}|\mathrm{B_i})$$
$$=\frac{1}{10}\sum_{\mathrm{i}=2}^{5}\frac{\mathrm{i}-1}{9}+\frac{1}{10}\sum_{\mathrm{i}=6}^{9}\frac{\mathrm{i}-2}{9}=\frac{16}{45}$$

例4.4 设一信号接收器在[0,1]时间上到达 n 个信号的概率为

$$p_n=\frac{\mu^n}{n!}e^{-\mu}\quad(\mu>0\text{常数})\quad n\in N_0=\{0,1,2,\cdots\}$$

一信号到达时，能被记录下来的概率是 0.4。设各信号到达时能否被记录相互独立，求该接收器在[0,1]上记录 l 个信号的概率，$l\in N_0$。

解: 记 A=接收器在[0,1]上记录 l 个信号，B_n=接收器在[0,1]上到达 n 个信号，$n\in N_0$。

$$A=A\Omega=\bigcup_{n=l}^{\infty}(B_nA),\quad P(B_n)=\frac{\mu^n}{n!}e^{-\mu},P(A|B_n)=C_n^l(0.4)^l(0.6)^{n-l}。$$

故

$$P(A)=\sum_{n=l}^{\infty}(\frac{\mu^n}{n!}e^{-\mu})C_n^l(0.4)^l(0.6)^{n-l}\quad\text{令}i=n-l$$
$$=\frac{(0.4\mu)^l}{l!}e^{-\mu}\sum_{i=0}^{\infty}\frac{(0.6\mu)^i}{i!}$$
$$=\frac{(0.4\mu)^l}{l!}e^{-0.4\mu}\qquad l\in N_0=\{0,1,2,\cdots\}$$

由以上几个例子可以看出，全概率公式的形式虽然很简单，但它所代表的“对样本空间进行适当分解”的思想却是十分重要，具有很高的技巧性。通过对样本空间的分解，我们可以将“较复杂的”事件转化与分解为若干不相容事件的并，从而使问题得以解决。这种事件分解的思想在后面章节仍会不断涉及，它贯穿概率论课程的始终，是概率论的一大特点。

对于全概率公式，还可以做以下两点推广：

推论 1 若$\{B_i,i\ge 1\}$是Ω的一个分解，且 $P(B_i)>0$，则

$$P(A)=\sum_{n=1}^{\infty}P(B_i)P(A|B_i)\qquad(4.4)$$

(证明提示：利用概率的可列可加性。与定理 4.3 证明类似)

推论 2 若 $\forall i \neq j, B_i B_j = \phi$，$B_i \in \Phi$ ，$A \subset \bigcup_{n=1}^{\infty} B_n$ ，则（4.4）亦成立。

全概率公式是概率论的最基本公式之一，它隐含了概率论中分析问题的基本思想与技巧，以后我们还要介绍它的各种变形与推广。

（3）贝叶斯公式（*Bayes* 公式）

由全概率公式，可知：$\mathrm{P}(A) = \sum_{i=1}^{n} \mathrm{P}(B_i)\mathrm{P}(A \mid B_i)$ 由乘法公式，得到：

$$\mathrm{P}(AB_i) = \mathrm{P}(A)\mathrm{P}(B_i \mid A) = \mathrm{P}(B_i)\mathrm{P}(A \mid B_i) \tag{4.5}$$

将（4.3）式代入（4.5）式，得到：$\sum_{i=1}^{n} P(B_i)P(A \mid B_i) \cdot P(B_i \mid A) = P(B_i)P(A \mid B_i)$

所以

$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{\sum_{i=1}^{n} P(B_i)P(A \mid B_i)} \tag{4.6}$$

定义 4.3 称(4.6)式为 *Bayes* 公式。

与全概率公式类似，*Bayes* 公式也可以推广到可列无穷的情况，即：

$$P(B_i \mid A) = \frac{P(B_i)P(A \mid B_i)}{\sum_{i=1}^{\infty} P(B_i)P(A \mid B_i)} \tag{4.7}$$

通过 *Bayes* 公式，使我们在已知 $P(B_i)$和 $P(A|B_i)$的情况下，可以计算 $P(B_i|A)$的值，我们称 $P(B_i|A)$为事件 B_i 的后验概率（即已知试验结果 A 以后 B_i 的概率）。相对地，称 $P(B_i)$为事件 B_i 的先验概率。

以下几个例子说明 *Bayes* 公式的应用。

例 4.5 某发报机每次随机的发“1”与“0”两种信号，其概率分别为 0.6 与 0.4，由于系统受随机干扰，其结果是：在发报机发“1”信号条件下，收报机收到“1”与“0”的条件概率为 0.8 与 0.2；在发“0”信号条件下，收到“1”与“0”的条件概率为“0.1”与“0.9”。试求在收报机收到“1”信号条件下，发报机是发“1”的条件概率？

解：设 A=发“1”信号，$\overline{A}$ =发“0”信号；B=收“1”信号，$\overline{B}$ =收“0”信号。

由题意有 $P(A) = 0.6, P(\overline{A}) = 0.4,\ P(B \mid A) = 0.8, P(\overline{B} \mid A) = 0.2, P(B \mid \overline{A}) = 0.1, P(\overline{B} \mid \overline{A}) = 0.9$ 。

由 *Bayes* 公式

$$P(A|B)=\frac{P(A)P(B|A)}{P(A)P(B|A)+P(\overline{A})P(B|\overline{A})}=\frac{0.6\times0.8}{0.6\times0.8+0.4\times0.1}=0.923。$$

此例中，$P(A)=0.6, P(\overline{A})=0.4$ 是根据以往发信号的统计规律得到的，相对于这个问题而言是先验概率，而 $P(A|B)$是根据本次得到的试验结果的信息下得到的后验概率。

例 4.6 对以往数据分析结果表明：当机器调整良好时，产品的合格率为 90%，而当机器发生某一故障时，其产品合格率仅为 30%；每天早上机器开动时，机器调整良好的概率为 75%。已知某日早上第一件产品是合格品，试求：机器调整良好的概率是多少？

解：A＝{产品合格}，B＝{机器调整良好}，则题目要求的是 $P(B|A)$；由题意：$P(A|B)$=90%，$P(\mathrm{A}|\overline{\mathrm{B}})=30\%$，$P(B)$=75%，所以

$$\mathrm{P}(B|A)=\frac{\mathrm{P}(B)\mathrm{P}(A|B)}{\mathrm{P}(B)\mathrm{P}(A|B)+\mathrm{P}(\overline{B})\mathrm{P}(A|\overline{B})}=\frac{0.75\times0.9}{0.75\times0.9+0.25\times0.3}=0.9$$

即，当生产出的第一件产品合格时，机器调整良好的概率为 90%。题目中 $P(B)$＝75%是根据以往的数据分析得到的，是先验概率；而 $P(B|A)$是在得到“生产了一件合格产品”这一信息条件下的修正概率，是后验概率。

（4）乘法公式、全概率公式在条件概率下的形式

在某事件 C 发生的条件下，乘法公式、全概率公式仍然成立，其形式如下：

乘法公式：设 $A,B,C\in\Phi$，若 $P(AC)>0$，则 $P(AB|C)=P(A|C)P(B|AC)$ (4.8)

全概率公式：$A\in\Phi$，若 $A\subset\bigcup_{i=1}^{\infty}B_i$，$B_i\in\Phi$，$\forall i\neq j, B_iB_j=\phi$，$\mathrm{P}(CB_i)>0$，则

$$P(A|C)=\sum_{i=1}^{\infty}P(B_i|C)P(A|CB_i) \tag{4.9}$$

上述形式的乘法公式和全概率公式，仍是无条件下相应公式的推广，其证明与后者类似；这里不再赘述，留给读者作为练习。

3 综合举例

在引入了条件概率的三个重要公式——乘法公式，全概率公式,*Bayes* 公式后，我们所解决的概率问题的范围比以前大大扩展了。为说明它们的综合应用，举例如下：

例 4.7 有两个口袋，一个装有 2 个黑球和 3 个红球；另一个装有 3 个黑球和 2 个红球；每次抛掷一枚匀称的硬币来决定从哪个袋中取球。每次抽球后均放回原口袋。若抽出的第一个球是黑球，求从同一口袋中抽得第二个球也是黑球的概率。

解：记 A_1＝从 1 号口袋(2 黑 3 红)取球；A_2＝从 2 号口袋(2 红 3 黑)取球；$B_1(B_2)$＝第1(2)个球是黑的。由题意，要求 $P(B_2|B_1)$。由条件概率定义：$P(\mathrm{B}_2|\mathrm{B}_1)=P(\mathrm{B}_1\mathrm{B}_2)/P(\mathrm{B}_1)$；再对事件 B_1 进行分解：$B_1=B_1(A_1\cap A_2)=A_1B_1\cap A_2B_1$；有：

$$P(\mathrm{B}_1)=P(\mathrm{A}_1\mathrm{B}_1)+P(\mathrm{A}_2\mathrm{B}_1)=P(\mathrm{A}_1)P(\mathrm{B}_1\mid \mathrm{A}_1)+P(\mathrm{A}_2)P(\mathrm{B}_1\mid \mathrm{A}_2)=\tfrac{1}{2}\cdot\tfrac{2}{5}+\tfrac{1}{2}\cdot\tfrac{3}{5}=\tfrac{1}{2}$$

类似地，

$$P(\mathrm{B}_1\mathrm{B}_2)=P(\mathrm{A}_1)P(\mathrm{B}_1\mathrm{B}_2\mid \mathrm{A}_1)+P(\mathrm{A}_2)P(\mathrm{B}_1\mathrm{B}_2\mid \mathrm{A}_2)=\tfrac{1}{2}\cdot(\tfrac{2}{5})^2+\tfrac{1}{2}\cdot(\tfrac{3}{5})^2=\tfrac{13}{50}$$

所以 $P(\mathrm{B}_2\mid \mathrm{B}_1)=\dfrac{P(\mathrm{B}_1\mathrm{B}_2)}{P(\mathrm{B}_1)}=\dfrac{13/50}{1/2}=\dfrac{13}{25}$。

与计算 $P(B_1)$类似可得：$P(B_2)=1/2<13/25$。这说明："第一次取到黑球"这一信息加大了第二次取得黑球的可能性，因为它加大了我们从含较多黑球的口袋取球的可能性。

推而广之，若已知前两次抽得的是黑球，则从同一口袋中第三次抽到的也是黑球的概率为：

$$P(B_3\mid B_1B_2)=\frac{P(B_1B_2B_3)}{P(B_1B_2)}=\frac{\tfrac{1}{2}\cdot(\tfrac{2}{5})^3+\tfrac{1}{2}\cdot(\tfrac{3}{5})^3}{\tfrac{1}{2}\cdot(\tfrac{2}{5})^2+\tfrac{1}{2}\cdot(\tfrac{3}{5})^2}=\frac{7}{13}.$$

一般的，若已知前 $n-1$ 次抽得的是黑球，则从同一口袋中第 n 次抽到的也是黑球的概率为：

$$P(B_n\mid B_1B_2\cdots B_{n-1})=\frac{P(B_1B_2\cdots B_n)}{P(B_1B_2\cdots B_{n-1})}=\frac{(\tfrac{2}{5})^n+(\tfrac{3}{5})^n}{(\tfrac{2}{5})^{n-1}+(\tfrac{3}{5})^{n-1}}。$$

当$n\to\infty$时，上述概率的极限为：

$$\lim_{n\to\infty}P(B_n\mid B_1B_2\cdots B_{n-1})=\lim_{n\to\infty}\frac{(\tfrac{2}{5})^n+(\tfrac{3}{5})^n}{(\tfrac{2}{5})^{n-1}+(\tfrac{3}{5})^{n-1}}=\frac{3}{5}=P(B_1\mid B_2)$$

例 4.8 里耶(*G．Polya*)坛子模型

有些人把这个问题当作传染病或地震的模型，认为某地越爆发疾病则越容易爆发该疾病。模型如下（红球代表爆发，黑球代表不爆发）：坛子中有 b 个黑球，r 个红球；现从中每次取一个，取后放回，并同时再放入 c 个与所取球同色的球。问：

（1）前二次取出的都是黑球的概率；

（2）第 n 次抽取的是黑球的概率。

解：令 B_i=第 i 次取出的是黑球，$i=1,2,\cdots$。

（1） 求 $P(B_1B_2)$：由乘法公式，有

$$P(\mathrm{B}_1\mathrm{B}_2)=P(\mathrm{B}_1)P(\mathrm{B}_2\mid \mathrm{B}_1)=\frac{b}{b+r}\cdot\frac{b+c}{b+r+c}。$$

（2）求 $P(B_n)$：

先考虑 $c=0$ 与 $c=-1$ 的特殊情形。这是我们已经熟悉的有放回抽样问题与无放回抽样，易知 $P(B_n)=\frac{b}{b+r}$；下面我们证明这个结论带有一般性。即对任意的 c，都有 $P(B_n)=\frac{b}{b+r}$。

考虑：在前 n 次抽取中，对任意指定顺序取 k 个黑球和$(n-k)$个红球的概率。每抽取

一次，球的总数加 c，每抽到一个黑(红)球，黑(红)球数加 c。均有：

$$\mu(\Omega)=C_{b+r}^{1}C_{b+r+c}^{1}\cdots C_{b+r+(n-1)c}^{1}\text{。}$$

μ(指定 k 个位置取黑球，其余(n-k)个位置取红球)= $C_{b}^{1}C_{b+c}^{1}\cdots C_{b+(k-1)c}^{1}\cdot C_{r}^{1}C_{r+c}^{1}\cdots C_{r+(n-k-1)c}^{1}$

P(指定顺序取 k 个黑球,(n-k)个红球)= $\dfrac{b(b+c)\cdots[b+(k-1)c]\cdot r(r+c)\cdots[r+(n-k-1)c]}{(b+r)(b+r+c)(b+r+2c)\cdots[b+r+(n-1)c]}$

记 A_k=前 n 次抽球中抽到 k 个黑球，k=0,1,2,···,n

则由全概率公式：（$\Omega=\bigcup_{k=0}^{n}A_k$ ）

$$\begin{aligned}P(B_{n+1})&=\sum_{k=0}^{n}P(A_k)P(B_{n+1}\mid A_k)\\&=\sum_{k=0}^{n}C_n^k\frac{b(b+c)\cdots[b+(k-1)c]\cdot r(r+c)\cdots[r+(n-k-1)c]}{(b+r)(b+r+c)(b+r+2c)\cdots(b+r+(n-1)c)}\cdot\frac{b+kc}{b+r+nc}\end{aligned}$$

若记 b′ =b+c，则

$$\begin{aligned}P(B_{n+1})&=\frac{b}{b+r}\sum_{k=0}^{n}C_n^k\frac{b'(b'+c)\cdots[b'+(k-1)c]r(r+c)\cdots[r+(n-k-1)c]}{(b'+r)(b'+r+c)\cdots[b'+r+(n-1)c]}\\&=\frac{b}{b+r}\sum_{k=0}^{n}\text{P(在}b'\text{个黑球，}r\text{个红球的坛子中，前}n\text{次取}k\text{个黑球}\\&=\frac{b}{b+r}\end{aligned}$$

由此可见，$P(B_n)$的概率与 n 及 c 均无关。不爆发传染病的概率是不随时间而转移的；也就是说，以前爆发与否对现在没影响。

§5 事件的独立性

1 两个事件的独立性

例 5.1 有 10 个球，6 白 4 红；现从中每次取出一个，取后放回；设 A_k=“第 k 次取出的是红球”，由古典概型的概率计算，可以得到：$P(A_1)=\frac{4}{10}$，$P(A_2)=\frac{4}{10}$，

$P(A_2\mid A_1)=\frac{4}{10}$，$P(A_2\mid\overline{A}_1)=\frac{4}{10}$， 即在 A_1 发生或不发生的条件下，A_2 发生的条件概率都等于 A_2 的无条件概率。这说明 A_1 的发生与否对 A_2 发生的概率没有影响。这时，我们可以设想 A_1、A_2 是彼此独立的，再由 $P(A_1A_2)=P(A_1|A_2)\cdot P(A_2)$，得：$P(A_1A_2)=P(A_1)\cdot P(A_2)$，由此引出两个事件独立的定义：

定义 5.1 若事件 A、B 满足：

$$P(AB)=P(A)\cdot P(B) \tag{5.1}$$

则称 A、B 相互独立。

对于独立性，有如下的命题：

命题 5.1 下列命题等价：（1）A 与 B 相互独立；（2）A 与 $\overline{B}$ 独立；（3）$\overline{A}$ 与 B 独立；（4）$\overline{A}$ 与 $\overline{B}$ 独立；（5）$P(A|B)=P(A)$；（6）$P(A\,|\,\overline{B})=P(A)$；（7）$P(A\,|\,\overline{B})=P(A), P(\overline{A}\,|\,\overline{B})=P(\overline{A})$。

证明：（1）□（2）。由 $A=AB\bigcup A\overline{B}$ 得：$P(A)=P(AB)+P(A\overline{B})$；由 A、B 相互独立，得：$P(AB)=P(A)P(B)$；则 $P(A\overline{B})=P(A)-P(AB)=P(A)-P(A)P(B)=P(A)(1-P(B))=P(A)P(\overline{B})$，故 $A,\overline{B}$ 相互独立。同理可证（2）□（3）□（4）□（1）。其余证明留给读者练习。

2 几个事件的独立性

（1）三事件的独立性

定义 5.2 如果三事件 A,B,C 满足：

$$\begin{cases} P(AB)=P(A)P(B) & (5.2)\\ P(AC)=P(A)P(C) & (5.3)\\ P(BC)=P(B)P(C) & (5.4)\\ P(ABC)=P(A)P(B)P(C) & (5.5)\end{cases}$$

则称 A,B,C 相互独立。

当事件 A,B,C 只满足：(5.2), (5.3), (5.4) 时，称事件 A,B,C 两两独立。

一般地，事件 A,B,C 两两独立时 $P(ABC)=P(A)P(B)P(C)$不一定成立，即事件 A,B,C 的两两独立性不能保证事件 A,B,C 的相互独立性，我们从下例可看到这点。

命题 5.2 下列命题等价：（1）A,B,C 独立；（2）$\overline{A},B,C$ 独立；（3）$A,\overline{B},C$ 独立；（4）$A,B,\overline{C}$ 独立；（5）$\overline{A},\overline{B},C$ 独立；（6）$\overline{A},B,\overline{C}$ 独立；（7）$A,\overline{B},\overline{C}$ 独立；（8）$\overline{A},\overline{B},\overline{C}$ 独立。

证明：类似于命题 5.1 的证明，从略。

作为练习，请读者证明若 A,B,C 独立，则 AB 与 C，$A\backslash B$ 与 C，$A\cap B$ 与 C 相互独立，一般的，有命题 5.3。

命题 5.3 设 A,B,C 独立，记$\Phi_1=\sigma(A,B)$，$\Phi_2=\sigma(C)$，任取 $A_1\in$ ☞$_1$, $A_2\in$ ☞$_2$，则 A_1,A_2 独立。

例 5.2 一个均匀的正四面体，第一面全染红色，第二面全染黄色，第三面全染蓝色，第四面染红黄蓝三色。在桌上将该四面体任意掷一次，考察和桌面接触的那一面上出现的颜色。记 A＝“出现红色”，B＝“出现黄色”，C＝“出现蓝色”。下面我们将指出这

三个事件两两独立， 但不相互独立；再记 A_i=“第 i 面接触桌面”，i=1,2,3,4；由题意 $P(A_i)$=1/4；又 $A = A_1 \cup A_4, B = A_2 \cup A_4, C = A_3 \cup A_4$，所以 $P(A)=P(B)=P(C)=1/2$；又 $AB=AC=BC=A_4$，故 $P(AB)=P(BC)=P(AC)=1/4$；可见事件 A、B、C 两两独立。

又 $P(ABC)=P(A_4)=1/4$，$P(A)P(B)P(C)=1/8$，所以 $P(ABC) \neq P(A)P(B)P(C)$，可见事件 A,B,C 不相互独立。

（2） n 个事件的独立性

由定义 5.2 推而广之，得到 n 个事件独立性的定义：

定义 5.3　n 个事件 $A_1, A_2, \cdots, A_n$，如果对任意的 $k(2 \le k \le n)$，有

$$P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k}) \tag{5.6}$$

其中 $i_1, i_2, \cdots, i_k$ 是满足 $1 \le i_1 < i_2 < \cdots < i_k \le n$ 的任意 k 个自然数，则称 $A_1, A_2, \cdots, A_n$ 相互独立。

注意：$P(A_{i_1}A_{i_2}\cdots A_{i_k}) = P(A_{i_1})P(A_{i_2})\cdots P(A_{i_k})$ 代表的等式个数为：

$$C_n^2 + C_n^3 + \cdots + C_n^n = (1+1)^n - C_n^1 - C_n^0 = 2^n - n - 1$$

还有如下命题：

命题 5.4 若 $A_1, A_2, \cdots, A_m, A_{m+1}, \cdots, A_n$ 相互独立 $(1 \le m < n)$，记 $\Phi_1=\sigma(A_k, 1 \le k \le m)$，

$\Phi_2=\sigma(A_k, m+1 \le k \le n)$，任取 $B_1 \in$ ☞$_1$, $B_2 \in$ ☞$_2$，则 B_1 与 B_2 相互独立。

例 5.3　甲、乙、丙三人向同一飞机射击，设各人是否命中相互独立，他们的命中率分别为 0.4、0.5、0.7。若只有一人射中飞机坠毁的概率为 0.2，若二人同时射中飞机坠毁的概率为 0.6，若三人同时射中，飞机一定坠毁；求飞机坠毁的概率。

解：飞机坠毁这一事件可看成由以下几个互不相容事件的并组成：因一个人击中而坠毁，因二人击中而坠毁，因三人击中而坠毁；故可依此思路进行样本空间的分解。

设 B＝飞机坠毁，A_i＝恰由 i 个人同时击中 $(i=0,1,2,3)$，则 $\Omega＝A_0＋A_1＋A_2＋A_3$

记 C_1＝甲射中，C_2＝乙射中，C_3＝丙射中，则

$A_0=\overline{C}_1\overline{C}_2\overline{C}_3$，$A_1=(C_1\overline{C}_2\overline{C}_3)\cup(\overline{C}_1C_2\overline{C}_3)\cup(\overline{C}_1\overline{C}_2C_3)$，

$A_2=(\overline{C}_1C_2C_3)\cup(C_1\overline{C}_2C_3)\cup(C_1C_2\overline{C}_3)$，$A_3=C_1C_2C_3$。

由已知数据及概率可加性及乘法公式可计算出：

$P(A_0) = 0.6 \times 0.5 \times 0.3 = 0.09$

$P(A_1) = 0.4 \times 0.5 \times 0.3 + 0.6 \times 0.5 \times 0.3 + 0.6 \times 0.5 \times 0.7 = 0.36, P(A_2) = 0.41, P(A_3) = 0.14$

由题意：$P(B|A_0)=0$，$P(B|A_1)=0.2$，$P(B|A_2)=0.6$，$P(B|A_3)=1$，故：

$$P(B)=\sum_{n=0}^{3}P(A_i)P(B|A_i)=0.09\times 0+0.36\times 0.2+0.41\times 0.6+0.14\times 1=0.458\text{。}$$

即飞机坠毁的概率为 0.458。

3 试验的独立性和独立重复试验序列

（1） 试验的独立性

记随机试验 $E^{(k)}$的样本空间为 $\Omega^{(k)}$，概率空间为($\Omega^{(k)}$，$\Phi^{(k)}$, $P^{(k)}$), ($1\le k\le n$)，两个试验 $E^{(1)}$ 与 $E^{(2)}$相互独立，粗略的说，即 $E^{(1)}$中的每个事件 $A^{(1)}\in\Phi^{(1)}$与 $E^{(2)}$中的任一事件 $A^{(2)}\in\Phi^{(2)}$相互独立。详述如下：

定义 5.3 记 n 个试验 $E^{(k)}$ 的复合试验为 $E\triangleq E^{(1)}\times E^{(2)}\cdots E^{(n)}\triangleq\mathop{\times}\limits_{i=1}^{n}E^{(i)}$，它对应的样本空间为$\Omega\triangleq\Omega^{(1)}\times\Omega^{(2)}\cdots\Omega^{(n)}\triangleq\mathop{\times}\limits_{i=1}^{n}\Omega^{(i)}$，相应的概率空间为(Ω，Φ, P)，其中：$\Phi\triangleq\mathop{\times}\limits_{i=1}^{n}\Phi_i$ P 是(Ω，Φ)上的概率测度，$\forall A^{(k)}\in\Phi^{(k)}$ (即 E 中仅与第 k 次试验 $E^{(k)}$有关的事件，($1\le k\le n$))如果

$$P(A^{(1)}A^{(2)}\cdots A^{(n)})=P(A^{(1)})P(A^{(2)})\cdots P(A^{(n)}), \tag{5.7}$$

则称试验 $E^{(1)},E^{(2)},\cdots,E^{(n)}$ 相互独立。

注：以上定义中的Φ 可以这样理解：$\Phi=\sigma\{A^{(1)}\times A^{(2)}\times\cdots\times A^{(n)}, A^{(k)}\in\Phi^{(k)}, 1\le k\le n\}$。

一般地，对于可列个试验 $E^{(1)},E^{(2)},\cdots,E^{(n)},\cdots$，若其中任意有限个相互独立，则称$\{E^{(n)},n\ge 1\}$为独立试验序列。

在这里我们用事件的独立性导出了试验的独立性定义。

（2） 独立重复试验序列

定义 5.4 若 $E^{(1)},E^{(2)},\cdots,E^{(n)}$相互独立，且 $E^{(i)}=E$，$\Omega^{(i)}=\Omega$，$\Omega=\mathop{\times}\limits_{i=1}^{n}\Omega^{(i)}=\Omega^n$；则称 $E^{(1)},E^{(2)},\cdots,E^{(n)}$为 n 次独立重复试验序列。若每个试验只关心一个事件 A 的发生与否（例如射击中的中与不中），则称之为 n 次贝努利(*Bernoulli*)试验。

直观地讲，n 次独立重复试验序列就是指独立地、重复地进行同一个试验 n 次；其中每次试验的每个结果不受其它试验结果的影响。

例5.4 设某人独立重复的射击 n 次，每次击中目标的概率为 p（$0\le p\le 1$），求

（1）给定的 k 次击中目标，而其余均未击中的概率（$0\le k\le n$）;

（2）恰好击中 k 次（$0\le k\le n$）；

（3）至少有一次击中。

解：令 B_k 为给定的 k 次击中目标，而其余均未击中；C_k 为恰好击中 k 次；D 为至少有一次击中。由独立性知

$$P(B_k)=p^k(1-p)^{n-k},P(C_k)=C_n^k p^k(1-p)^{n-k},P(D)=1-(1-p)^n 。$$

例 5.5 通讯中常采取重复发送信号的办法来减少在接收中可能发生的错误。假定发报机只发 0 和 1 两种信号，接收时发生错误（0 收为 1 或 1 收为 0）的概率为 0.05。为减小错误，采取每一信号连发三次，接收时按数目多的来判定信号，求在此情形下判错一个信号的概率。

解：每个信号的判定包含连续三次信号的接收，这三次信号的接收可视为相互独立的贝努利试验。设 A_i=“有 i 次接收错误”（i=1,2,3），B=“判定信号发生错误”。

由题意 $P(B)=P(A_2)+P(A_3)$ 又：$P(A_i)=C_3^i(0.05)^i(1-0.05)^{3-i}$

所以$\mathrm{P(B)}=\mathrm{C}_3^2(0.05)^2(1-0.05)+\mathrm{C}_3^3(0.05)^3=0.00725$，因此，判错的概率大大减小了。

4 事件的条件独立性

定义 5.5 若事件 A、C 及 B 满足：

$$P(AC|B)=P(A|B)\cdot P(C|B), \tag{5.8}$$

则称 A、C 关于事件 B 条件独立。

命题 5.5 下列命题等价：

（1）A，C 关于 B 条件独立；

（2）$P(A|BC)=P(A|B)$；

（3）$P(A|B\overline{C})=P(A|B)$；

（4）$P(C|AB)=P(C|B)$

（5）$P(C|\overline{A}B)=P(C|B)$

（6）A，$\overline{C}$ 关于 B 条件独立。

证明：略。（有兴趣的读者可作为练习，自己证之，并考虑是否还有其它等价命题）

（1）（2）（3）说明 A，C 关于 B 条件独立，其直观意义是：在 B 事件发生的条件下，C 发生与否不影响 A 发生的条件概率。

注意：A，C 关于 B 条件独立，一般不能推出 A，C 独立；反之亦然。

例 5.5 考虑一个电子游戏机，它有两个状态“1”和“2”。记 B=“1”，$\overline{B}$=“2”，A_k 为第 k 次操作游戏机出现成功事件。已知 $P(B)=0.3$，$P(A_k|B)=0.15$，$P(A_k|\overline{B})=0.20, k=1,2$；且各次出现成功与否关于系统状态“1”和“2”是条件独立的，即 $P(A_1A_2|B)=P(A_1|B)P(A_2|B)$，$P(A_1A_2|\overline{B})=P(A_1|\overline{B})P(A_2|\overline{B})$。

试求 $P(A_1\cap A_2)$，并证 A_1,A_2 不独立。

解：$P(A_k)=P(A_k\,|\,B)P(B)+P(A_k\,|\,\overline{B})P(\overline{B})=0.15\times 0.30+0.20\times 0.70=0.185$

$$\begin{aligned}P(A_1A_2)&=P(A_1A_2\,|\,B)\,P(B)+P(A_1A_2\,|\,\overline{B})\,P(\overline{B})\\&=P(A_1\,|\,B)\,P(A_2\,|\,B)\,P(B)+P(A_1\,|\,\overline{B})\,P(A_2\,|\,\overline{B})\,P(\overline{B})\\&=0.0348\end{aligned}$$

故 $P(A_1\cup A_2)=P(A_1)+P(A_2)-P(A_1A_2)=0.335$

$$P(A_1\,|\,A_2)=\frac{P(A_1A_2)}{P(A_2)}=0.1878\neq 0.1850=P(A_1)$$

所以 A_1,A_2 不独立。这说明 A_1,A_2 关于 B 和关于 $\overline{B}$ 条件独立，不能保证 A_1,A_2 独立。

练习题

1.1 从一副扑克牌的 13 张黑桃中，一张接一张地有放回的抽取 3 张，求：

（1） 没有同号的概率；

（2） 有同号的概率；

（3） 最多只有两张同号的概率。

1.2 某人有 5 把钥匙，但忘了开房门的是哪一把，逐把试开，问：

（1）恰好第三次打开房门锁的概率是多少？

（2）三次内打开的概率是多少？

（3）如 5 把内有 2 把房门钥匙，三次内打开的概率是多少？

1.3 从 6 双同样规格的手套中任取四只，问其中恰有一双的概率是多少？

1.4 某人写了 n 张信笺与 n 个信封，将信笺随便乱装入信封内（每个信封装一张信笺），

问无一信匹配的概率是多少？

1.5 两人约定某日晚 7 点至 8 点在某地会面，试求一人要等另一人半小时以上的概率。

1.6 指出下列各式中哪些成立，哪些不成立：

(1)$A\cup B=(A\overline{B})\cup B$;

(2)$\overline{A}B=A\cup B$;

(3)$\overline{A\cup B}\,\overline{C}=\overline{A}\,\overline{B}\,\overline{C}$;

(4)$(AB)(A\overline{B})=\phi$;

（5） 若 $A\subset B$, 则 $A=AB$。

1.7 盒中有 12 个乒乓球，其中 9 个是新的。第一次比赛时从中任取 3 个来用，比赛后放回盒中。第二次比赛时再从盒中任取 3 个，求第二次取出的球都是新球的概率。又：如第二次取出的球都是新球，求第一次取到的都是新球的概率。

1.8 设某昆虫生产 k 个卵的概率为：

$$p_k=\frac{\lambda^k}{k!}e^{-\lambda}\qquad(\lambda>0)\,;$$

又设一个虫卵能孵化为昆虫的概率为 p 。若卵能否孵化为虫是相互独立的。问此昆虫的下一代有 m 条的概率是多少？

1.9 某产品 40 件，其中有次品 3 件，现从中任取 2 件，求其中至少有一件次品的概率。

1.10 袋中有红、黄、白色球各一个，每次任取一个，有放回地抽三次，求下列事件的概率：*A*=“三个都是红的” = “全红”，*B*=“颜色全同”，*C*=“颜色全不同”，*D*=“颜色不全同”，*E*=“无黄”，*F*=“无红且无黄”。

1.11 从一副扑克牌中，任意抽出两张，问都是黑桃的概率是多少？

1.12 从一副扑克牌中，任取 13 张，求各点彼此不同的概率。

1.13 设有 n 个人，问此 n 个人彼此有不同生日的概率是多少？其中至少有两个人有相同生日的概率是多少？（设每人出生在一年 365 日中的每一天的机会是均等的）。

1.14 分别印有 1，2，…，10 的 10 个球，装在一个袋中，现从中任取 3 个，问：大小在中间的号码恰为 5 的概率是多少？

1.15 由盛有号码 1，2，…，N 的球的箱子中，有放回地抽了 n 次球（每次一个），依次记下其号码，试求：（1）这些号码按严格上升次序排列的概率；（2）这些号码按上升（不一定严格上升）次序排列的概率。

1.16 10 个白球 10 个黑球被随机分为 10 组，每组两球，求每组中恰有一白球一黑球的概率。

1.17 某码头只能容纳一只船。现预知某日将独立来到两只船，且在 24 小时内各时刻来到的机会均等。如果它们需要停靠的时间分别为 3 小时与 4 小时，试求有一船要在江中等待的概率。

1.18 在一张打上方格的纸上投一枚直径为 1 的硬币，问：方格要多小才能使硬币与网格线不相交的概率小于 0.01？

1.19 袋中有 2 个红球，3 个黄球，5 个白球，现有放回地抽取了八次（每次一个）。问：恰抽到 3 红 3 黄 2 白的概率是多少？

1.20 在一个袋子里有 n 张带 1 至 n 号码的票，我们一张一张地随机取出票来(不放回），问：至少有一次所取的票号码与次序数相一致的概率为多少？

1.21 某人带有两盒火柴，每盒 n 根。每次用时他在两盒中随机抓一盒，从中取出一根。问：遇到一盒空而另一盒剩 r 根的概率为多少（$r\leqslant n$）？

1.22 设样本空间 $\Omega=\{\omega_1,\omega_2,\omega_3,\omega_4\}, A=\{\omega_1\}, B=\{\omega_1,\omega_2\}$,集类✌ $=\{A,B\}$。（1）试 用 A,B,Ω 的运算关系表示下列各事件：A，B 中至少有一个发生；A 不发生，B 发生；（2）写出由集类A生成的 σ—域 σ(✌) 中的所有元素。

1.23 设 $\Omega=R=(-\infty,+\infty)$，任取 $a\in R$ 固定，记

$$A=\{x:x\leq a\}, A_n=\{x:x<a+1/n\},$$

$$B=\{x:x<a\}, B_n=\{x:x\leq a-1/n\}, n\in N=\{1,2,\cdots\}。$$

证明：$A=\bigcap\limits_n A_n$，$B=\bigcup\limits_n B_n$。

1.24 设 $\Omega=R$,记：

✌ $_1=\{(-\infty,a],a\in R\}$,✌ $_2=\{(a,b],a,b\in R\}$,✌ $_3=\{(a,b),a,b\in R\}$，

✌ $_4=\{[a,b),a,b\in R\}$,✌ $_5=\{[a,b],a,b\in R\}$。

(1) 试用集类A_2 的元素的运算表示A_1 中的元素，用A_1 中的元素的运算表示A_2 中的元素；

(2) 证明：$A_1\subset\sigma(A_2)$，$A_2\subset\sigma(A_1)$；

(3) 证明：✍🖬 σ☎✌☎) $=\sigma$(✌ $_2$) $=\sigma$(✌ $_3$) $=\sigma$(✌ $_4$) $=\sigma$(✌ $_5$)。

注：称B为一维 *Borel* σ—域，B中元素称一维 *Borel* 集。

1.25 设(Ω, Φ, P)=([0,1], B[0,1], P)是[0,1]上的 *Borel* 概率空间。记 B=[0,1]上有理数全体，$\overline{B}$=[0,1]上无理数全体。试用概率公理化定义 与性质求：

$$P\left((1/5,1/4]\right), P\left((1/5,1/4)\right), P\left((1/5,1/4]\cup(1/3,1/2)\right), P(B)及P(\overline{B})。$$

1.26 给定概率空间 $(\Omega$, ☞,$P)$，设 $A,B\in$ ☞ 且 $P(A)=P(B)=1$，试用概率性质求：

$$P(A\overline{B}), P(AB), P(A\cup B), P(A\cup\overline{B})。$$

1.27 设甲和乙两同学接连要做 3 次随机试验，每人每次试验均有 2 种可能结果：成功和失败。记 A_k 与 B_k 分别表甲、乙第 k 次试验成功（1≤k≤3）已知：

$P(A_1)=P(B_1)=0.6$，$P(A_2\mid\overline{A_1})=0.8$，$P(B_2\mid\overline{B_1})=0.05$，$P(A_3\mid\overline{A_1}\,\overline{A_2})=0.98$

$P(B_3\mid\overline{B_1}\,\overline{B_2})=0$。试分别求甲成 功（至少有一次成 功）的概率和乙成功的概率。

1.28（条件概率全概率公式）设 $A,B_n,C\in$ ☞$,n\geq 1,(B_k,k\geq 1)$ 两两互不相容 $A\subset\bigcup_{k=1}^{\infty}B_k$ 且 $P(B_kC)\neq 0$。试证：$P(A\mid C)=\sum_{k=1}^{\infty}P(B_k\mid C)P(A\mid B_kC)$。

1.29 设 A，B 为两事件，A，B 独立，证明下列命题等价：（1）A 与 $\overline{B}$ 独立；（2）$\overline{A}$ 与 $\overline{B}$ 独立；（3）$P(A\mid B)=P(A)$；（4）$P(A\mid\overline{B})=P(A)$；（5）$P(B\mid A)=P(B)$；（6）$P(B\mid\overline{A})=P(B)$。

1.30 设三事件 A,B,C 相互独立，证明：$A\cup B, A\setminus B, AB, \overline{A}\overline{B}, \overline{A}B$ 均与 C 独立；$\overline{A},\overline{B},\overline{C}$ 相互独立。

1.31 设 A_1,A_2,A_3,A_4 是 Ω 的划分，$P(A_k)=1/4, 1\leq k\leq 4, A=A_1\cup A_4$，$B=A_2\cup A_4$，$C=A_3\cup A_4$。证明：$A$，$B$，$C$ 两两独立，但它们不独立。

1.32 称事件 A,C 关于 B 条件独立,若 $P(AC\mid B)=P(A\mid B)P(C\mid B)$。证此命题与下列命题等价：（1）$A,\overline{C}$ 关于 B 条件独立 ；（2）$\overline{A},C$ 关于 B 条件独立 ；

(3) $\overline{A},\overline{C}$ 关于 B 条件独立；(4) $P(A\mid BC)=P(A\mid B)$；

(5) $P(A\mid B\overline{C})=P(A\mid B)$；（6）$P(\overline{A}\mid BC)=P(\overline{A}\mid B)$；

(7) $P(\overline{A}\mid B\overline{C})=P(\overline{A}\mid B)$；(8) $P(A\mid BC)=P(A\mid B\overline{C})$ 。

试列举其他的等价命题。

1.33 甲、乙两人独立的射击同一目标，记 A=“甲击中”，B=“乙击中”，已知 $P(A)$=0.6，$P(B)$=0.7，现甲乙各射击一次 ，若已知目标被击中，求这是被甲击中的条件概率。

1.34 袋中有 10 个球，6 白 4 红，随机地一个个取出（不放回），求在第 4 次首次取到红球的概率。

1.35 试证命题 2.1。

1.36 试证概率性质(*vii*)中的(2.13)式。

第二章　随机变量及其分布

在前一章的学习中，我们已经讨论了随机数学中一些最基本的概念和一些最简单的概型中随机事件的概率求解方法，并给出了概率的公理化定义。这样，我们在遇到一些简单的具体的随机事件时，可以先判断其概型，然后利用已有知识，对其概型特性进行分析，求出概率。但在实际工作中，我们遇到的概率问题往往是错综复杂、千变万化的，并不都是我们熟知的、简单的概型。为更全面研究随机实验结果以揭示随机事件中包含的客观存在的统计规律，我们将随机实验的结果量化，这样就引入了“随机变量”的概念。

引入随机变量这个概念后，可以用随机变量表示随机事件，以便将不同的随机事件的概率及相互关系进行统一处理，而且有助于研究随机事件的本质规律。更为重要的是它极大的丰富了研究对象与范围，它可以解决涉及人类生活中更为广泛的理论与应用的问题。随机变量是随机数学中最重要的基本概念之一。

§ 1.　随机变量

1　随机变量

先给出一些直观例子：

例 1.1　投篮时会出现两种情况：ω_1=“投中”，ω_0=“未投中”。于是$\Omega=\{\omega_0,\omega_1\}$。我们若用数字 1 代表投中，用 0 代表未投中。这样便将实验结果量化了，同时引入了一个变量 X，就是

$$X=X(\omega)=\begin{cases}1, & \omega=\omega_1,\\ 0, & \omega=\omega_0.\end{cases}$$

$X(\omega)$ 随实验结果 ω不同而取不同的值，由于实验结果 $\omega\in\{\omega_0,\ \omega_1\}$出现是随机的，因此 $X(\omega)$的取值也是随机的，$X(\omega)$是随机变量。

例 1.2　在某学校中，随机的抽取一人 ω，用对应的卡片 ω登记他的身高 $L(\omega)$, 将这些卡片放在一起，若从中随机地取出一张卡片，那么 $L(\omega)$ 的取值随实验结果 ω的不同而不同，因而 $L(\omega)$取值也是随机的，$L(\omega)$也是随机变量。

定义 1.1　设 E 为随机试验，样本空间 $\Omega=\{\omega\}$为有限集或可列集，若对每个样本点 $\omega\in\Omega$，都有一个实数 $X(\omega)$与其对应，即 $X(\omega)$是定义在 Ω上取值在 R 上的单值实值函数（映射），简记 $X(\omega)$： $\Omega\to R$，称 $X(\omega)$为**随机变量**(*random variable*，简记作 *r.v.*)。

可以用随机变量 X 描述事件，例如在例 1.1 中随机变量 X 取值 1，记 $\{\omega:X(\omega)=1\}$ 或简记为$\{X=1\}$，就表示事件{投篮命中}。

例 1.3　某射手进行射击训练，他每次射中目标的概率是 p，且各次射击是否击中相互独立。如果射手只射击了 n 次，记这 n 次射击中命中目标的次数 X，那么 X 就是一个随机变量。如果他开始射击，直到第一次命中目标后停止射击，记他首次命中目标时的射击次数为 Y，Y 也是一个随机变量。

前面对随机变量只在样本空间为有限集或可列集时给出粗略的定义，其确切的定义需要用到波雷尔(*Borel*)σ–域的知识。下面我们先来说一下什么是波雷尔σ–域（有关波雷尔集的知识在第一章中已有介绍）。

2　波雷尔σ–域与概率空间

(1)　设 $\Omega=R$，考虑 R 中右半闭的无穷区间 $(-\infty,a]$，全体这样的区间构成的集类A，即$A=\{(-\infty,a];a\in R\}$，称由A生成的σ–域σ(A)为波雷尔σ–域。通常记为B。B中的集称为波雷尔集。

设 $g(x)$是定义在 R^1 上的单值实函数，如果对任意 $a\in R$，有$(x:g(x)\leqslant a)\in$B，则称 $g(x)$ 为波雷尔可测函数，或B–可测函数。

(2)　设 $\Omega=R^n$，n 为任一正整数，R^n 中的右半闭无穷区间记为$\prod_{i=1}^{n}(-\infty,a_i]$。它是一个 n 维点集。全体这样的区间构成 R^n 中的集类A：

$$A=\{\prod_{i=1}^{n}(-\infty,a_i],\ a_i\in R^n,i=1,2,\dots,n\}$$

称由A生成的σ–域为 n 维波雷尔σ–域, 通常记为$B^n\triangleq\sigma(A)$，B^n 中的集称为 n 维波雷尔集。

设 $g(x_1,x_2,\dots,x_n)$是定义在 R^n 上的单值实函数，如果对任意实数 a，有$\{g(x_1,x_2,\dots,x_n)\leqslant a\}\in B^n$，则称 $g(x_1,x_2,\dots,x_n)$为 n 元波雷尔可测函数，或B^n–可测函数。

对于一般的概率空间(Ω,Φ,P)，同样可以给出定义在 Ω上取值在 R 上的实函数 $X(\omega)$，但是注意到引出 $X(\omega)$的一个重要目的是为了用它统一表示我们关心的事件，并能够定义其概率。为此，要求 $X(\omega)$具有某些良好的性质，即对 $\forall a\in R$，有 $\{\omega:X(\omega)\leq a\}\in$ ☞，我们称满足此性质的 $X(\omega)$为(Ω,Φ,P) 上的随机变量。

随机变量的确切定义如下：

定义 1.2　设(Ω,Φ,P)是一个概率空间，$X(\omega)$是定义在 $\Omega=(\omega)$上的一个实值函数。如果对$\forall a\in R$，有

$$\{\omega:X(\omega)\leq a\}\in ☞ \qquad (1.1)$$

则称 $X(\omega)$为(Ω,Φ,P)上的随机变量。

几点说明：

1) $(\omega:X(\omega)\leqslant a)$是指所有满足 $X(\omega)\leqslant a$ 的样本点 ω 的集合。定义要求

$(\omega : X \le A) \in$ ☞，即（ω：$X(\omega) \leqslant a$）是(Ω，Φ，P)的一个事件，因而可定义它的概率。

2）定义中 ω 为自变量，为了书写方便简记（ω：$X(\omega) \leqslant a$）=（$X \leqslant a$）=（$X \in (-\infty, a]$）以下把 $X(\omega)$ 记为 X。一般随机变量符号用大写字母 X,Y,Z 等表示。

3） $X(\omega)$满足（1.1），则容易验证：$\forall\, a,b \in R$ 且 $a \leqslant b$，$(X>a)$,$(X<a)$, $(X=a)$, $(a<X\leqslant b)$, $(a\leqslant X<b)$, $(a<X<b)$, $(a\leqslant X\leqslant b) \in \Phi$。（见练习题）

例 1.4 设 A 为任一事件，$A \in \Phi$，$I_A(\omega)$为事件 A 的示性函数。由于 ω满足：

Comment:

$$\{I_A(\omega) \le x\} = \begin{cases} \phi \in \text{☞} & \text{当} \quad x < 0 \quad \text{时} \\ \overline{A} \in \text{☞} & \text{当} \quad 0 \le x < 1 \quad \text{时} \\ \Omega \in \text{☞} & \text{当} \quad x \ge 1 \quad \text{时} \end{cases}$$

所以 $I_A(\omega)$是一个随机变量。

Comment:

例 1.5 设 $\Omega=\{\omega_1, \omega_2, \omega_3\}$

$X(\omega)$	ω_1	ω_2	ω_3
P	1/6	1/3	1/2

解：设 $X(\omega_1)=2$， $X(\omega_2)=4$，$X(\omega_3)=5$。则上表可变成下表：

$X(\omega)=i$	2	4	5
$P(\omega:X(\omega)=i)$	1/6	1/3	1/2

由上表可见：

$$P(X(\omega) \le 1) = 0 \qquad P(X(\omega) \le 3) = P(X(\omega) = 2) = \frac{1}{6}$$

$$P(X(\omega) \le 4) = P(X(\omega) = 2) + P(X(\omega) = 4) = \frac{1}{6} + \frac{1}{3} = \frac{1}{2}$$

$$P(X(\omega) \le 5) = P(X(\omega) = 2) + P(X(\omega) = 4) + P(X(\omega) = 5) = \frac{1}{6} + \frac{1}{3} + \frac{1}{2} = 1$$

$$P(X(\omega) \le 10) = 1$$

例 1.6 若（Ω，Φ）中的$\Phi=\{\phi, A, \overline{A}, \Omega\}$。$A \subset \Phi$而 $A_1 \notin$ ☞，容易验证 A_1 的示性函数 $I_{A_1}(\omega)$ 使 $(\mathrm{I}_{A_1} \le \frac{1}{2}) \notin$ ☞。故 $I_{A_1}(\omega)$ 对$\Phi=\{\phi, A, \overline{A}, \Omega\}$ 而言不满足 $r.v.$的定义。

Comment:

例 1.7 （Ω，Φ），设 $\{B_k\}(0 \le k < \infty)$ 是 Ω的一个划分，即 $B_k B_l = \phi \quad (k \ne l)$；$\bigcup_{k=1}^{\infty} B_k = \Omega$ 且 $B_k \in$ ☞$, 0 \le k < \infty$，定义 $X(\omega) = \sum_{k=1}^{\infty} x_k I_{B_k}(\omega)$，则容易验证 $X(\omega)$是 $r.v.$。

注意：$\forall a \in \mathbf{R}$，$\{\omega : X(\omega) \le a\} \in$ ☞。可见随机变量的概念，是对已给的(Ω, Φ, P)中的σ–域Φ 而言的。对于只含有穷或可列多个样本点的 Ω，通常取Φ 是由 Ω中的全体子集构成的σ–域，则定义在 Ω上的实函数均是随机变量。

在第一章我们定义了什么是由集类生成的σ−域，引出随机变量以后，我们自然要问，什么是由随机变量生成的σ−域？

定义 1.3 设 X 是一 $r.v.$，则称$\sigma(X)=\sigma(\omega: X(\omega)\le a, \forall a\in R)$为由 X 生成的σ−域。

例如：由示性函数 I_A 生成的σ−域为：$\sigma(I_A)=\{\phi, A, \overline{A}, \Omega\}$。

附：以下命题初学者看时只需了解结论即可，证明可先略去不看，这不影响本书的学习。

命题 1.1* 对任意 $x\in R$，$(X(\omega)\leqslant x)\in\Phi$ 成立的充要条件是：对任意 $A\in$B，有$(X(\omega)\in A)\in\Phi$。

证明：充分性，设$(X(\omega)\in A)\in\Phi$对任一 $A\in$B成立。取 $A=(-\infty,x]\in$B 即得：$(X(\omega)\leqslant x)\in\Phi$，对 $x\in R$ 成立。

必要性：令 $☹=\{A: A\in ✍, (X(\omega)\in A)\in ☞\}$，即Λ 是由B中一切使$(X(\omega)\in A)\in\Phi$成立的 A 构成的集类。由Λ定义，有 $☹\subset ✍$；这样为证Λ= B，只需证 $☹\supset ✍$，为此只需证Λ是包含 $✌=\{(-\infty,x], x\in R\}$ 的σ−域。

下证Λ 是σ−域：事实上，$R\in ✍$，$(X(\omega)\in R)=\Omega\in\Phi$，故 $R\in\Lambda$；其次，若 $A\in\Lambda$，则 $A\in$B。且$(X(\omega)\in A)\in\Phi$；因B是σ−域，所以$\overline{A}=\mathbf{R}-A\in ✍$；再注意Φ是 σ−域，得 $\{X(\omega)\in\overline{A}\}=\overline{\{X(\omega)\in A\}}\in ☞$，所以 $\overline{A}$ 也满足上述两个条件，从而 $\overline{A}\in\Lambda$，知Λ对余封闭。又设 $A_i\in\Lambda, i=1,2,\cdots$，即 $A_i\in ✍$，$\bigcup_{i=1}^{\infty}\{X(\omega)\in A_i\}\in ☞$；又B及Φ是σ−域，得$\bigcup_{i=1}^{\infty}A_i\in ✍$，$\{X(\omega)\in\bigcup_{i=1}^{\infty}A_i\}=\bigcup_{i=1}^{\infty}\{X(\omega)\in A_i\}\in ☞$，故$\bigcup_{i=1}^{\infty}A_i\in\Lambda$，所以Λ是包含 R 的σ−域。再证Λ包含 $✌=\{(-\infty,x], x\in R\}$，因为 $\forall x\in R$，$\{X(\omega)\in(-\infty,x]\}=\{X(\omega)\le x\}\in ☞$，所以$(-\infty,x]=A$ 满足：$A\in ✍$且$\{X(\omega)\in A\}\in ☞$，所以$(-\infty,x]\in\Lambda$。这样便证明了Λ是包含A=$\{(-\infty,x]; x\in R\}$的σ−域。于是 $☹\supset\sigma(✌)=✍$，得Λ=B。
综上引理得证。

常用的随机变量有两类：一类是离散型随机变量，另一类是连续型随机变量。下面两节将介绍这两类随机变量。

§ 2　离散型随机变量

1　定义

定义 2.1　如果随机变量 $X(\omega)$所有可能取值是有限个或可列多个，则称 $X(\omega)$为离散型随机变量（*discrete random variable*）简写作 *d.r.v.*。

例如§1 中的 X 和例 1.3 中的 X 与 Y 都是离散型随机变量。刻画 *d.r.v.X* 的特性，不仅要知道它的可能取值，更要知道它取各值的概率。

定义 2.2　设离散型随机变量 (*d.r.v.*) $X(\omega)$所有可能取的值为 x_k ,$k\in N=\{1,2,\ldots\}$,$X(\omega)$取各个值的概率为：

$$P\{X=x_k\}=p_k\ ,\quad k\in N \tag{2.1}$$

称 $P\{X=x_k\}=p_k\ (k\in N)$为 *d.r.v* $X(\omega)$的概率分布或分布律。

r.v $X(\omega)$的概率分布反映了它取各种可能值的概率分配，包含了它的全部概率信息。由概率定义，易知 p_k 满足如下两个条件：

$$\begin{aligned}&(1)\qquad p_k\geq 0,\quad k\in N\ \ (\text{非负性})\\&(2)\qquad \sum_{k=1}^{\infty}p_k=1\qquad (\text{规一性})\end{aligned}\tag{2.2}$$

可以这样理解分布律，将质量 1 分布在各个质点上，x_k 点处的质量为 p_k 。

如图 2-1

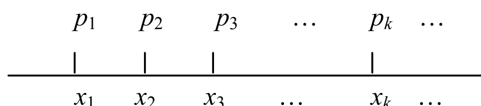


图 2-1

分布律也可以用表格形式来表示：

X	x_1	x_2	x_3	…	x_k
P_k	p_1	p_2	p_3	…	p_k

2　几种重要的离散型随机变量的概率分布

（1）　二点分布

定义 2.3　若 *r.v.* X 只取 1 和 0 两个值，且 $P(X=1)=p$，$P(X=0)=1-p$，$(0\leq p\leq 1)$，则称 *r.v.X* 服从参数为 p 的二点分布。简记为：$X\sim B(1,p)$。

若事件 A 发生 *r.v.X* 取 1，否则 X 取 0；即

$$\forall A\in\mathscr{F},\quad X(\omega)=I_A(\omega)=\begin{cases}1 & \omega\in A\\0 & \omega\notin A\end{cases}$$

则 $P(A)=P\{I_A(\omega)=1\}=p$，$P(\overline{A})=P\{I_A(\omega)=0\}=1-p$；可见示性函数可用来描述某随机事件发生与否。

例 2.1　200 件产品，190 件是合格的，10 件是不合格的，现从中任取一件，若规定

$$X=\begin{cases}1 & 若取得合格产品\\ 0 & 若取得不合格产品\end{cases}$$

则 X 服从参数为 0.95 的二点分布。

二点分布是最简单的一种分布类型，它可描述一切只有或只关心两种可能结果的随机事件。比如产品合格与不合格，新生婴儿是男是女，比赛中的胜与负，电信号的正与负，种子是否发芽等等。

（2）　二项分布（*Binomial distribution*）

以 X 表示 n 重贝努利试验中 A 发生的次数，易知 X 是一个随机变量，其可能取值为 $0,1,2,\ldots,n$。由于各次试验相互独立，故在 n 次试验中 A 发生 k 次的概率 $P(X=k)=C_n^k p^k(1-p)^{n-k},\qquad k=0,1,2,\ldots,n$

定义 2.4　若 $r.v.X$ 满足:

$$P(X=k)=C_n^k p^k(1-p)^{n-k},\qquad k=0,1,2,\ldots,n \tag{2.3}$$

则称 $r.v.X$ 服从参数为（n,p）的二项分布，又称为贝努利分布（*Bernoulli* 分布)，简记 $X\sim B(n,p)$。

根据上面的定义，显然有 $P(X=k)\geq 0$(非负性)；$\sum_{k=1}^{n}C_n^k p^k q^{n-k}=(p+q)^n=1$（规一性)。特别地，当 n=1 时，二项分布化为：$P(X=k)=p^k q^{1-k},\quad k=0,1$，这就是二点分布。可见，二点分布是二项分布的一个特例。

例 2.2　某人进行射击训练，每次射中的概率时 0.02，独立射击 400 次，求至少击中 1 次的概率。

解：　将每次射击看作一次独立试验，则整个试验可看作一个 400 次的贝努利试验。设击中的次数为 X，则 $X\sim B(400,0.02)$。X 的分布率为：

$$P(X=k)=C_{400}^k 0.02^k 0.98^{400-k},\qquad k=0,1,2,\ldots,400.$$

则所求概率为 ：$P(X\geq 1)=1-P(X=0)=1-0.98^{400}\approx 0.9997$

这个例子的实际意义十分有趣。这个射手每次命中的概率只有 0.02，绝不是个天才，但他坚持射击 400 次，则击中目标的概率近似为 1，几乎成为必然事件。这正说明了，“只要功夫深，铁杵磨成针”。不要因为成功的希望小而放弃，只要我们锲而不舍的努力，就一定会到达理想的彼岸。这一道简单的概率题中蕴含着深刻的哲理。

上例中，直接计算 $P(X=k)$相当麻烦，下面我们介绍一个当 n 相当大，p 很小时二项分布概率的近似公式，这就是二项分布的泊松逼近。

定理 2.1 (泊松(*Poisson*)定理)　设 $\lambda>0$ 是一个常数，n 是任一正整数，设 $p_n=\lambda/n$，则对任一固定的非负整数 k，有

$$\lim_{n\to\infty} C_n^k {p_n}^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!} \tag{2.4}$$

证明：$\because \quad p_n = \dfrac{\lambda}{n}$

$$\therefore \quad C_n^k {p_n}^k (1-p_n)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k \left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$= \frac{\lambda^k}{k!}\left[1\cdot(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})\right](1-\frac{\lambda}{n})^n (1-\frac{\lambda}{n})^{-k}$$

对任一固定的 k，当 $n\to\infty$ 时

$$(1-\frac{\lambda}{n})^n \to e^{-\lambda}, \qquad (1-\frac{\lambda}{n})^{-k} \to 1, \quad (1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n}) \to 1,$$

故：$\displaystyle \lim_{n\to\infty} C_n^k {p_n}^k (1-p_n)^{n-k} = \frac{\lambda^k e^{-\lambda}}{k!}$　　证毕.

由于 $np_n=\lambda$（常数），所以 $n\to\infty$ 时 $p_n\to 0$。因此当 n 很大，p_n 很小时有以下近似公式：

$$C_n^k {p_n}^k (1-p_n)^{n-k} \approx \frac{\lambda^k e^{-\lambda}}{k!}, \qquad 其中\lambda = np_n$$

在例 2.1 中，$\lambda=np=400\times0.02=8$

故

$$P(X=0)=\frac{\lambda^0 e^{-\lambda}}{0!}=e^{-8} \qquad P(X=1)=\frac{\lambda^1 e^{-\lambda}}{1!}=8e^{-8}$$

$$P(X\ge 2)=1-P(X=0)-P(X=1)=1-e^{-8}-8e^{-8}\approx 0.997$$

例 2.3　为了保证设备正常工作，需配备适量的维修工人。现有同类型设备 300 台独立工作，发生故障的概率都是 0.01。在通常情况下一台设备的故障可由一个人来处理。问至少需要多少人才能保证当设备发生故障，不能及时维修的概率小于 0.01?

解：　设至少需配备 N 人，同一时刻发生故障的设备有 X 台，那么 $X\sim B(300,0.01)$。依题目,要求的 N，应使　$P\{X\geqslant N\}<0.01$　　　　　　　　　　（*）

根据泊松定理（$\lambda=np_n=3$），故：$P\{X>N\}=\displaystyle\sum_{k=N+1}^{\infty}\frac{3^k e^{-3}}{k!}$。所以式(*)化为：

$\displaystyle\sum_{k=N+1}^{\infty}\frac{3^k e^{-3}}{k!}<0.01$，通过查表，可知满足上式的 $N>8$，因此至少需 9 人。

（3）　**泊松（*poisson*）分布**

定义 2.5　若 *r.v.X* 满足：

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!},\qquad k\in N_0=\{0,1,2,\cdots\} \tag{2.5}$$

Comment:

其中 λ>0 是常数，则称 *X* 服从参数为 λ 的泊松分布，记为 $X\sim Po(\lambda)$。

由定义易知：

$$P\{X=k\}=\frac{\lambda^k e^{-\lambda}}{k!}\geq 0, k\in N_0\text{；}\quad \sum_{k=0}^{\infty}P\{X=k\}=\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}=e^{-\lambda}\sum_{k=0}^{\infty}\frac{\lambda^k}{k!}=e^{-\lambda}e^{\lambda}=1$$

可见 $P\{X=k\}$满足概率分布的条件。

在信息通讯、计算机网络、生物学、医学、工业管理及公用事业的排队等问题中，泊松分布是常见的。例如一“服务台”在给定时间上到达的“顾客”数，容器内微生物数，玉器上的疵点数，交换台的电话呼唤次数等，大多服从泊松分布。由泊松分布引出的泊松信号流是随机过程的一个重要分支，对它的详细阐述见第六章。

例 2.4　自 1875 年至 1955 年中某 63 年间，某市夏季（5~9 月）共发生暴雨 180 次。每年夏季共有 *n*=31+30+31+31+30=153 天。每次暴雨以 1 天计算，每天发生暴雨的概率则为 *p*=180/(63×153)，这个值很小。但 *n* =153 很大。应用泊松分布，在一个夏季发生 *k* 次暴雨的概率 p_k 为：

$$p_k=\frac{2.9^k e^{-2.9}}{k!}\qquad (\lambda=np_k=153\times\frac{180}{63\times 153}\approx 2.9)$$

我们可通过下表来看计算值为实际观测之间相符合的情况。例如，发生 2 次暴雨的实际有 14 个夏天，计算值为 14.8 个。基本相符。

暴雨次数	0	2	4	6	8
实际年数	4	14	10	2	1
理论年数	3.5	14.8	10	2.9	0.42

（4）　**几何分布**

定义 2.6　若 *d.r.v.X* 取值正整数，且满足：

$$P\{X=k\}=(1-p)^{k-1}p,\qquad k=1,2,\ldots\quad (0\leq p\leq 1) \tag{2.6}$$

则称 *r.v.X* 服从参数为 *p* 的几何分布，记为 $X\sim G(p)$。

在 *Bernoulli* 独立重复试验中，若事件 *A* 在一次试验中出现的概率为 *p*，即 *P*(*A*)=*p* 记 *X* 为首次发生 *A* 的试验次数，则 $X\sim G(p)$。

几何分布具有一个极特殊的性质——无记忆性。我们假设在前 *n* 次试验中事件 *A* 均未发生，则可证明，在此条件下，从第 *n*+1 次试验起到首次 *A* 发生的第 *X* 次试验，仍服从几何分布，与 *n* 无关，即：$\forall n\in N$，有

$$P(X=n+k \mid X>n)=P(X=k)=(1-p)^k p,\quad k=1,2,\cdots \tag{2.7}$$

这就是几何分布的无记忆性。

可以证明，在离散型分布中，只有几何分布才具有无记忆性。（证明略，见习题）

3　离散型随机变量(*d.r.v.*)的分解与表示

给定(Ω, Φ, P)，若$A_n\in\Phi$，$n\in N$两两不交，且$\sum_{n=1}^{\infty}A_n=\Omega$,即$\{A_n, n\geq 1\}$是$\Omega$的一个分解，$a_n\in \mathrm{R}$，$n\in N$；易证：$X(\omega)=\sum_{n=1}^{\infty}a_n I_{A_n}(\omega)$，$\omega\in\Omega$ 为离散型随机变量。反之，对于任一*d.r.v.X*，若其分布律为：$P(X=x_k), k\in N$，因$B_k \hat{=} (X=x_k)\in ☞, k\in N$，则可以表示为

$$X(\omega)=\sum_{k=1}^{\infty}x_k I_{B_k}(\omega)，\omega\in\Omega \tag{2.8}$$

这说明对于任一*d.r.v.*，总可以分解为互不交的事件的示性函数的迭加。

定义 2.7　形如(2.8)式定义的函数,称为(Ω, Φ)上的简单函数.

例 2.5　二项分布可用示性函数表示。

若随机变量$X(\omega)\sim B(n,p)$，$B_k=\{X=k\}$,（$0\leqslant k\leqslant n$），则

$$X(\omega)=\sum_{k=0}^{n}k\cdot I_{\{X=k\}}(\omega)=\sum_{k=0}^{n}k\cdot I_{B_k}(\omega)。$$

用示性函数的线性组合表示*d.r.v.*这种分解的思想与方法在随机数学，特别是现代随机数学中起着巨大的作用。

§3　连续型随机变量

1　定义

定义 3.1 若对于*r.v.X*，存在一定义在R上的非负函数$f(x)$，使对$\forall L\in$ ☝，满足

$$P(X\in L)=\int_{x\in L}f(x)dx \tag{3.1}$$

则称X为连续型随机变量，简记：*c.r.v.X*；其中函数$f(x)$称为X的概率密度函数（*probability density function*,简记为*p.d.f.*），简称**概率密度**。

注意：由定义可看出，概率密度函数不是唯一的。因为若$f(x)$是*r.v.X*的*p.d.f.*，如在有限个点上或可列个点上改动它的值为另一非负数，记为$g(x)$，则两者在同一L上的积

分值相同，因而 $g(x)$亦是 X 的 $p.d.f.$。

可以证明定义 3.1 与下列定义 3.1a 等价。

定义 3.1*a* 若对于 $r.v.X$，存在一定义在 R 上的非负函数 $f(x)$，使对$\forall a\in R$，满足：

$$P(X \le a) = \int_{-\infty}^{a} f(x)dx$$

则称 X 为 $c.r.v.$，$f(x)$为 X 的 $p.d.f.$。

由定义易知，概率密度具有下列性质：

(1)　$f(x) \ge 0$　　(非负性)

(2)　$\int_{-\infty}^{+\infty} f(x)dx = 1$　　(规一性)

(3)　$P(a < X \le b) = \int_a^b f(x)dx$　　$(a \le b)$

(4)　如果$r.v.X(\omega)$有概率密度$f(x)$，则$P\{\omega: X(\omega)=a\}=0,(\ \forall a \in R)$。

(5)　若$f(x)$在x处连续，则：$f(x) = \underset{h\to 0}{Lim}\dfrac{P(x < X \le x+h)}{h}$　　（*）

即 $P(x < X \le x+h) = f(x)h + \circ(h)$，其中$\circ(h)$是 h 的高阶无穷小，即 $\underset{h\to 0}{Lim}\dfrac{o(h)}{h} = 0$。

性质（4）的证明：

$$P\{\omega: X(\omega)=a) = \underset{n\to+\infty}{Lim} P(a-\frac{1}{n} < X \le a+\frac{1}{n}) = \underset{n\to+\infty}{Lim}\int_{a-\frac{1}{n}}^{a+\frac{1}{n}} f(x)dx = 0$$

可见，概率为 0 的事件不一定是不可能事件。

有些随机变量的概率密度函数直接求解较为困难，此时可通过求微小区域上的概率的主要部分（见（*）式）来求 $f(x)$，这种求解概率密度的方法称作微元法。它是一种重要的方法，利用它可以使一些复杂问题得到简单明了的解决，读者在以后的学习中将会逐步地体会到这一点。

例 3.1　设随机变量 X 具有概率密度 $f(x) = ke^{-3x}I_{(x\ge 0)}$，试确定常数 k，并求 $P(X > 0.1)$。

解：由$\int_{-\infty}^{\infty} f(x) = 1$，解得 k=3，于是 X 的概率密度为：$f(x) = 3e^{-3x}I_{(x\ge 0)}$，所以

$P\{X > 0.1\} = \int_{0.1}^{\infty} f(x)dx = \int_{0.1}^{\infty} 3e^{-3x}dx = 0.7408$。

2　几种连续型随机变量的分布

（1） 均匀分布（*Uniform distribution*）

定义 3.3 若连续型随机变量 X 具有概率密度：

$$f(x)=\begin{cases}\frac{1}{b-a} & x\in[a,b] \quad (a<b)\\ 0 & \text{其它}\end{cases} \tag{3.2}$$

则称 X 在区间 $[a,b]$ 上服从均匀分布，记作 $X\sim U[a,b]$。

由定义可知，在区间$[a,b]$上服从均匀分布的随机变量 X，落在$[a,b]$ 的任一子区间的概率只依赖于子区间的长度而与子区间的位置无关。事实上，对于任一长度为 l 的子区间 $(c,c+l)$，$a\le c<c+l\le b$，有： $P\{c<X\le c+l\}=\int_c^{c+l}\frac{1}{b-a}dx=\frac{l}{b-a}$

记 $F(x)\triangleq P(X\le x)=\int_{-\infty}^{x}f(t)dt$ 可得： $F(x)=\begin{cases}0 & x<a\\ \frac{x-a}{b-a} & a\le x\le b\\ 1 & x>b\end{cases}$

$f(x)$,$F(x)$ 的图像分别如图 3-1、3-2 所示。

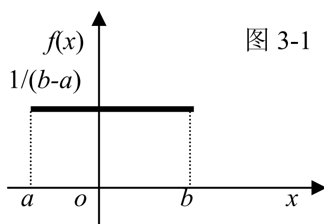


图 3-1

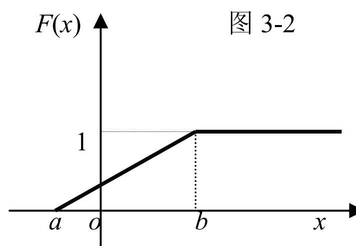


图 3-2

例 3.2 设电阻 R 是一随机变量，均匀分布在 900 Ω~1100 Ω，求 R 的概率密度及 R 落在 950 Ω~1050 Ω的概率。

解：由题意，R 的概率密度为： $f(r)=\begin{cases}\frac{1}{1100-900} & 900<x<1100\\ 0 & \text{其它}\end{cases}$

$$\therefore \quad P\{950\le R\le 1050\}=\int_{950}^{1050}\frac{1}{200}dr=0.5$$

（2） 正态分布（*Normal distribution*）

定义 3.4 若连续型随机变量 X 的概率密度为：

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad -\infty<x<+\infty \tag{3.3}$$

其中 μ, σ为常数且σ>0，则称 X 服从参数为（μ, σ^2）的正态分布或高斯(*Gauss*)分布，记为 $X \sim N(\mu, \sigma^2)$。

$f(x)$ 的图像如下所示：

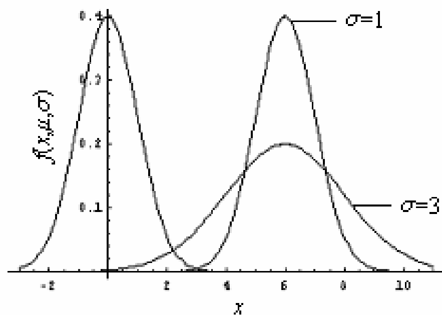


图 3—3

$f(x)$具有以下性质：

（1）　关于 $x=\mu$ 对称。这表明对于任意 $h>0$，有

$$P\{\mu-h<X\le\mu\}=P\{\mu<X\le\mu+h\}。$$

（2）　当 $x=\mu$ 时，$f(x)=f(\mu)=\dfrac{1}{\sqrt{2\pi}\sigma}$,取得最大值。$x\to\pm\infty$，$f(x)\to 0$，曲线以 x 轴为渐近线。对于同样长度的区间，当区间离 μ 越远，X 落在这个区间上的概率越小。

（3）　$f(x)$在 $x=\mu\pm\sigma$处有拐点。

（4）　σ越大，$f(x)$的图像越“胖”，表明取值较分散；σ越小，$f(x)$的图像越“瘦”，表明取值较集中。

（5）　$f(x)$存在 n 阶连续导数。

（6）　$\int_{-\infty}^{+\infty} f(x)dx=1$（规一性）。

证明：只证性质(6)。其它性质的证明留给读者作为练习。

记 $I=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$　　即证$I=1$；因 I 非负，只须证 $I^2=1$。

$$I^2=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(s-\mu)^2}{2\sigma^2}}ds=\frac{1}{2\pi\sigma^2}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{(t-\mu)^2+(s-\mu)^2}{2\sigma^2}}dtds$$

令 $x=\frac{t-\mu}{\sigma}$， $y=\frac{s-\mu}{\sigma}$，又令 $x=\rho\sin\theta$，$y=\rho\cos\theta$，$\rho>0$，$0\leqslant\theta<2\pi$

则： $I^2=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}e^{-\frac{x^2+y^2}{2}}dxdy=\frac{1}{2\pi}\int_0^{2\pi}\int_0^{+\infty}e^{-\frac{\rho^2}{2}}\rho d\rho d\theta=1$ 得证。

特别地，若 $X\sim N(0,1^2)$ 即 X 服从参数 $\mu=0,\sigma=1$ 的正态分布，则称 X 为标准正态分布。称

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-x^2/2} \tag{3.3.1}$$

为标准正态概率密度函数。

称

$$\Phi(x)=P(X\leq x)=\int_{-\infty}^{x}\varphi(x)dx=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-t^2/2}dt \tag{3.3.2}$$

为标准正态分布函数。对于 $\Phi(x)$ 的值，可通过查表得到（见附表 1）。易知 $\Phi(-x)=1-\Phi(x)$ 。

例 3.3 X 服从标准正态分布，求 $P\{|X|\leqslant1\}$，$P\{|X|\leqslant2\}$，$P\{|X|\leqslant3\}$。

解: $P\{|X|\leq1\}=P\{-1\leq x\leq1\}=\Phi(1)-\Phi(-1)=\Phi(1)-[1-\Phi(1)]=2\Phi(1)-1=0.6826$

同样，$P\{|X|\leqslant2\}=\Phi(2)-\Phi(-2)=0.9545$；$P\{|X|\leqslant3\}=\Phi(3)-\Phi(-3)=0.9973$。

一般地，若 $X\sim N(\mu,\sigma^2)$, 则只要通过一个适当的线性变换就能将它化成标准正态分布。此过程也称作对随机变量 X 进行标准化。

引理 若 $X\sim N(\mu,\sigma^2)$，则 $Z=\frac{X-\mu}{\sigma}\sim N(0,1^2)$

证明： 由 $Z=\frac{X-\mu}{\sigma}$，故

$$P\{Z\leq x\}=P\{X\leq\mu+\sigma x\}=\frac{1}{\sqrt{2\pi}\sigma}\int_{-\infty}^{\mu+\sigma x}e^{-\frac{(t-\mu)^2}{2\sigma^2}}dt\overset{令\frac{t-\mu}{\sigma}=u}{=}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du=\Phi(x)$$

由此 $Z=\frac{X-\mu}{\sigma}\sim N(0,1^2)$ 证毕。

这样，就可以利用标准正态分布来表示一般的正态分布函数。若 $X\sim N(\mu,\sigma^2)$，对于任意区间 $(a,b]$有：

$$P\{a<X\leq b\}=\Phi(\frac{b-\mu}{\sigma})-\Phi(\frac{a-\mu}{\sigma}) \tag{3.3.3}$$

查表即可得。

对于标准正态分布随机变量，引入 α 分位点的定义。

定义 3.5　设 X~$N(0,1^2)$若 Z_α满足条件 $P\{X > Z_\alpha\} = \alpha$　$(0 < \alpha < 1)$，则称点 Z_α为标准正态分布上的 α 分位点。

在自然现象和社会现象中，大量的随机变量都服从或近似服从正态分布。例如，海洋波浪的高度，一个地区成年男子的身高，生产条件不变的许多产品的某些量度都服从正态分布。正如二项分布主要产生于贝努利试验，泊松分布主要产生于泊松流，正态分布早期来源于测量误差分析。在概率论中和数理统计的理论研究和实际应用中，正态分布都占有重要地位。

（3）　指数分布（*Exponential distribution*）

定义 3.6　若连续型随机变量 X 的概率密度函数为：

$$f(x) = \lambda e^{-\lambda x} I_{(x\geq 0)}, \qquad (3.4)$$

其中 $\lambda>0$ 为常数，则称 X 服从参数为 λ 的指数分布，记作 X~ $Ex(\lambda)$。

指数分布的两个重要性质：

性质 1 (无记忆性)：对 $\forall s,t > 0$　有$P\{X > t\} = P\{X > s + t \mid X > s\}$。

证明：左= $P\{X > t\} = 1 - P\{X \leq t\} = e^{-\lambda t}$

$$右=\frac{P\{X > s, X > t + s\}}{P\{X > s\}} = \frac{P\{X > t + s\}}{P\{X > s\}} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda s}$$

$$\therefore \quad P\{X > t\} = P\{X > s + t \mid X > s\}$$

性质 2（无老化性），即瞬时失效率为常数，且 X 服从指数分布是 X 具有无老化性的充要条件。

证明：充分性：设 X~ $Ex(\lambda)$ 为工作寿命。

$\underset{h\to\infty}{Lim}\frac{P\{t < X < t + h \mid X > t\}}{h} = \frac{f(t)}{P(X > t)}, \forall t \geq 0$，即为 t 时刻的瞬时失效率。

$\forall t > 0, h > 0$

$$\underset{h\to 0}{Lim}\frac{P\{t < X < t + h \mid X > t\}}{h} = \underset{h\to 0}{Lim}\frac{P\{t < X < t + h\}}{P\{X > t\}h} = \underset{h\to 0}{Lim}\frac{\lambda e^{-\lambda t}}{e^{-\lambda t}} = \lambda$$

$\therefore$　$X \sim Ex(\lambda)$时，具有无老化性质。

必要性：记 $F(t) = P(X \leq t)$，因 X 具有无老化性质，故瞬时失效率 $F'(t)/(1 - F(t)) = \lambda$，

其中 λ 为常数。令 $F(0)=0$，有 $d(1 - F(t))/(1 - F(t)) = -\lambda dt$，

$F(t)=(1-e^{-\lambda t})I_{(t\geq 0)}=\int_0^t \lambda e^{-\lambda u}du$，即 $X\sim Ex(\lambda)$。证毕。

利用指数分布的这两条性质，有时可以简捷地解决一些涉及指数分布的问题。

（4）贝塔分布（*Beta distribution*）

定义 3.7 若 *r.v.X* 的 *p.d.f.*为：

$$f(x)=\frac{1}{B(\alpha,\beta)}\cdot x^{\alpha-1}(1-x)^{\beta-1}\cdot I_{(0\leq x\leq 1)} \quad (\alpha>0,\beta>0) \tag{3.5}$$

其中 $\mathrm{B}(\alpha,\beta)=\int_0^1 u^{\alpha-1}(1-u)^{\beta-1}du$ 称为贝塔函数，称 X 服从参数为 (α,β) 的贝塔分布；记为 $X\sim \mathrm{B}e(\alpha,\beta)$。

注意当 $\alpha=1,\beta=1$ 时，贝塔分布就变成[0，1]上的均匀分布。

（5）伽玛分布（*Gamma distribution*）

定义 3.8 若 *r.v.X* 的 *p.d.f.*为：

$$f(x)=\frac{1}{\Gamma(\alpha)}\alpha^{\gamma}x^{\gamma-1}e^{-\alpha x}I_{(x\geq 0)} \quad （\alpha>0,\gamma>0） \tag{3.6}$$

其中 $\Gamma(\alpha)=\int_0^{\infty}u^{\alpha-1}e^{-u}du$ 是伽玛函数，称 X 服从参数为 (α,γ) 的伽玛分布；记为 $X\sim \mathrm{G}(\alpha,\gamma)$。

注意当 $\gamma=1$ 时，就化为指数分布。

注：*Gamma* 函数有以下性质：$\Gamma(\nu)=(\nu-1)\Gamma(\nu-1)$；$\Gamma(1)=1$；$\Gamma(\frac{1}{2})=\sqrt{\pi}$；当 $\nu=n\in N$ 时，$\Gamma(n)=(n-1)!$；$\Gamma(\alpha)$ 与 $B(\alpha,\beta)$的关系为：$\mathrm{B}(\alpha,\beta)=\frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$。

§4 随机变量的分布函数

1 概率分布函数（*Probability distribution function*）

以上我们介绍了离散型和连续型两类随机变量，又分别介绍了针对它们的分布律和密度函数两种方式来描述随机变量的概率分布特性。其实，我们可以把二者统一起来，即可以引入所谓概率分布函数的概念，来统一表达一般随机变量的概率特性。

定义 4.1 设 $X(\omega)$是概率空间(Ω,Φ,P)上的随机变量，$\forall u\in\mathbf{R}$，称

$$F_X(u)=P(X(\omega)\leqslant u) \tag{4.1}$$

为 X 的概率分布函数，简称为 $d.f.$

说明：

（1）　概率分布函数 $F_X(u)$是定义在 R 上取值在[0，1]上的函数，它表示事件$(X(\omega)\leqslant u)$的概率，即 $F_X(u)$等于 X 落在（—∞，u]上的概率。当区间的右端点 μ 发生变化时，$P(X(\omega)\le\mu)$ 也发生相应地改变。因此，对给定的 X，$P(X\in(-\infty,u])$ 是区间$(-\infty$，$u]$右端点 u 的函数。

（2）　$r.v.X$ 的概率分布函数包含了它的全部概率特性的信息。（见下页注 4.1）

（3）　由第一章可知给定(Ω,Φ,P)，P 是定义于Φ上的集函数。这样为使 $P(X(\omega)\leqslant u)$有意义，要求 $\forall u\in R,(X\le u)\in$ ☞，这正是 $r.v.X$ 定义中要此限制的理由。

（4）　在不引起误解时，常将 X 省去，简记为 $F(x)$。

定理 4.1　概率分布函数 $F_X(x)$ $(x\in R)$具有下列性质：

(1)　单调不减性，即若 $b>a$，则 $F_X(b)\geqslant F_X(a)$

（2）　右连续：$\mathop{Lim}\limits_{b\to a^+}F_X(b)=F_X(a)$。

（3）　令 $F_X(-\infty)=\mathop{Lim}\limits_{x\to-\infty}F_X(x)$，$F_X(\infty)=\mathop{Lim}\limits_{x\to\infty}F_X(x)$，则 $F_X(-\infty)=0$，$F_X(\infty)=1$。

注：若定义 $F_X(x)=P(X<x)$，则 $F_X(x)$是左连续的。

证明：（1）如果 $b>a$，则 $(X(\omega)\leqslant b)=(X(\omega)\leqslant a)\cup(a<X(\omega)\leqslant b)$；等式右边两个事件不相容，故 $F_X(b)=P(X(\omega)\leqslant b)=P(X(\omega)\leqslant a)+(a<X(\omega)\leqslant b)\geqslant P(X(\omega)\leqslant a)=F_X(a)$

（2）由于 $F_X(x)$的单调不减性，故存在右极限 $F_X(a+0)=\mathop{Lim}\limits_{b\to a+}F_X(b)$；由于单调性，为证 $F_X(a+0)=F_X(a)$,只需对某一列$\{b_n\}$，$b_1>b_2>\ldots>b_n>a$，$b_n\to a$,证明当 $b_n\to a$ 时，有 $\mathop{Lim}\limits_{n\to\infty}F_X(b_n)=F_X(a)$ 即可。令 $A_n=(a<X(\omega)\leqslant b_n)$，显然 $A_n\supset A_{n+1}$，$\bigcap_{n=1}^{\infty}A_n=\phi$，于是由$(X(\omega)\leqslant b_n)=(X(\omega)\leqslant a)\cup(a<X(\omega)\leqslant b_n)$及第一章连续性定理，得

$$\mathop{\mathrm{Lim}}\limits_{n\to\infty}F_X(b_n)-F_X(a)=\mathop{Lim}\limits_{n\to\infty}[F_X(b_n)-F_X(a)]=\mathop{Lim}\limits_{n\to\infty}P(A_n)=P(\phi)=0$$

则$\mathop{Lim}\limits_{n\to\infty}F_X(b_n)=F_X(a)$，即 $\mathop{Lim}\limits_{b\to a^+}F_X(b)=F_X(a)$。

（3）令 $C_n=(X(\omega)\leqslant -n)$，则 $C_n\supset C_{n+1}$，$\bigcap_{n=1}^{\infty}C_n=\phi$，由连续性定理：$\mathop{Lim}\limits_{n\to\infty}F(-n)=\mathop{Lim}\limits_{n\to\infty}P(C_n)=P(\phi)=0$，则 $F_X(-\infty)=0$；同理可证：$F_X(\infty)=1$。

定义 4.2　满足上述（1）、（2）、（3）的函数 $F(x)$称为分布函数。

注 4.1：由此易知，任一 $r.v$ 的概率分布函数必是分布函数；同时，由测度论中已证其逆命题成立，即任一分布函数必是某一概率空间上一 $r.v$ 的概率分布函数，故对 $r.v$ 的概率分布函数，简称为分布函数。同时说明 $r.v$ 的分布函数包含了它的全部概率特性。

由定理 4.1 ，我们可以用 $F_X(x)$来表达一些重要事件的概率：

令$F_X(b-0)=\underset{a\to b^-}{Lim}F_X(a)$,

1) $P(a<X\le b)=F_X(b)-F_X(a)$; (4.2)

2) $P(X=b)=F_X(b)-F_X(b-0)$;

3) $P(X<b)=F_X(b-0)$;

4) $P(X>b)=1-F_X(b)$;

5) $P(X\ge b)=1-F_X(b-0)$.

$F_X(x)=P(X\le x)$，还可改写为：$F_X(x)=P(X\in(-\infty,x])$，这说明作为点 $x\in R$ 的函数 $F_X(x)$可看作是区间 $(-\infty,x]$上的函数。由第一章证明：

$\forall x\in R,(X\le x)\in$ ☞ $\Leftrightarrow \forall B\in$ 👌$,(X\in B)\in$ ☞ ,

故对任一 $A\in$B，$P(X(\omega)\in A)$都有定义。

例 4.1　已知 X 的概率分布律如下，求其分布函数 $F_X(x)$

X	-2	0	2
P	q^2	$2pq$	P^2

其中 $p+q=1$

解：当 $x<-2$ 时，$F_X(x)=P(X\le x)=P(X<-2)=0$，

当 $x\in[-2，0)$时，$F_X(x)=P(X\le x)=P(X=-2)=q^2$，

当 $x\in[0，2)$时，$F_X(x)=P(X\le x)=P(X=-2)+P(X=0)=q^2+2pq$，

当 $x\geqslant 2$ 时，$F_X(x)=P(X\le x)=P(X=-2)+P(X=0)+P(X=2)=1$。

$F_X(x)$ 的图像如图 4-1。

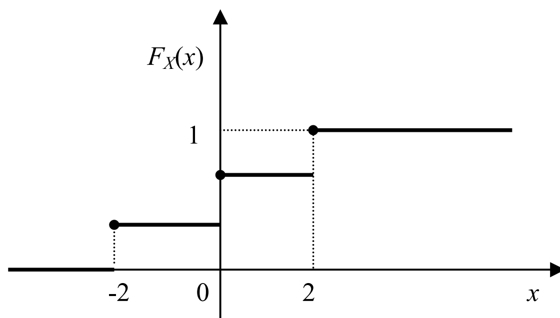


图 4-1

若$d.r.v.X$的分布律为$P(X=x_k)=p_k, k\in N, p_k\ge 0, \sum\limits_k p_k=1$,容易验证:$\forall a\in R$,有:

$$F_X(a)=P(X\le a)=\sum_{k:x_k\le a}P(X=x_k)=\sum_{k:x_k\le a}p_k \tag{4.3}$$

对于连续型 $r.v.$的 $p.d.f.$与其分布函数有如下关系:

命题4.1:若$r.v.X$具有$p.d.f, f_X(\mu), \mu\in R$,则$\forall x\in R$有

$$F_X(x)=P(X\le x)=\int_{-\infty}^{x}f_X(t)dt \tag{4.4}$$

证明:由定义(3.1)式,取$L=(-\infty,x)\in\beta$,即得(4.4)。

从概率密度函数的等价定义可知有如下命题:

命题4.2:设$r.v.X$的分布函数为$F_X(x)$,若存在定义在R上的$f(u), u\in R$,且对$\forall x\in R$,满足(4.4)式,则$f(u)$是X的$p.d.f$.

证明:从略.

例 4.3　$X\sim N(2, 1^2)$，令 $Y(\omega)=X(\omega)\vee 0 \hat{=} \max(X(\omega), 0)$，求 $Y(\omega)$的分布函数。

解：由题意知 $Y\geqslant 0$　当 $u<0$ 时，$F_Y(u)=P(Y\le u)=0$，当 $u\geqslant 0$ 时

$$F_Y(u)=P(Y\le u)=\Phi(\frac{u-2}{1})=\Phi(u-2)$$

注:$P(Y=0)=P(X\leqslant 0)=\Phi(-2)$,可见 Y 既不是离散型 $r.v$，也不是连续型 $r.v.$

若给定一个 $r.v.X$ 的分布函数 $F(x)$，也就给定了一个 $r.v.X$ 的所有的概率特性。事实上，若两个 $r.v.X,Y$,它们的分布函数相同,即 $\forall a\in R, P(X\le a)=P(Y\le a)$，则由本章命题 1.1

可以证明：$\forall L \in \mathrm{B}$，有 $P(X\in L)=P(Y\in L)$，这表明 *r.v.*的分布函数决定了它的所有概率特性。

§5　条件分布函数与条件密度函数

在第一章，我们讨论了条件概率，及一些相应特性，这一节，我们将在条件概率和分布函数的基础上，给出条件分布函数及条件密度函数的概念。

定义 5.1　给定(Ω, Φ, P)，设 *r.v.X* 及事件 $B\in\Phi$ ，$P(B)>0$，$\forall x \in R$，称

$$F_{X|B}(x) \triangleq P(\omega : X \le x \mid B) \tag{5.1}$$

为 X 关于事件 B 的条件（概率）分布函数。若存在非负函数 $f_{X|B}(x)$，使得：

$$\forall L \in \mathscr{B}: \int_{x\in L} f_{X|B}(x)dx = P(\omega : X(\omega) \in L \mid B) \tag{5.2}$$

则称 $f_{X|B}(x)$为 X 关于事件 B 的条件（概率）密度函数(*Conditional Prob. density function* 简称为 *c.p.d.f.*)。

例 5.1　$X\sim N(2，1^2)$，$B=\{\omega: X\geqslant 0\}$，求 X 在 B 条件下的条件分布函数及 *c.p.d.f.*。

解：　当 $x\leqslant 0$ 时：$P(X \le x \mid X \ge 0) = 0$；　当 $x>0$ 时：

$$P(X \le x \mid X \ge 0) = \frac{P(X \le x, X \ge 0)}{P(X \ge 0)} = \frac{P(0 \le X \le x)}{P(X \ge 0)} = (\Phi(x-2) - \Phi(0-2))/(1-\Phi(0-2))$$

$$= \frac{\Phi(x-2) - \Phi(-2)}{\Phi(2)}$$

当 $x<0$ 时,$F'_{X|X\ge 0}(x) = 0$；当 $x>0$ 时,$F'_{X|X\ge 0}(x) = (0.9772)^{-1}(2\pi)^{-\frac{1}{2}}\exp(-(x-2)^2/2)$；

X 在 $X\geqslant 0$ 条件下的 *c.p.d.f.*，可以取为：

$$f_{X|B}(x) = \begin{cases} 0 & x \le 0 \\ (0.9772)^{-1}(2\pi)^{-\frac{1}{2}}\exp(-(x-2)^2/2) & x > 0 \end{cases}$$

注意：上例 $F_{X|X\ge 0}(x)$ 在 $x=0$ 的导数不存在。

注：若一分布函数 $F(x)$在 R 上连续,且仅在可列点上导数不存在，则它的 *p.d.f.* $f(x)$ 可取为：在 $F(x)$导数存在点上取 $f(x) = F'(x)$；在导数不存在的点上任取 $f(x)= a\geq 0$ 即可。

若 *r.v.*$X\sim N(2，1^2)$，设事件 A 为 $(X\geqslant 0)$，定义 *r.v.Y*，$Y \triangleq X \cdot I_A(\omega)$。读者可再分析一下 Y 的分布函数（注意 Y 为一般性 *r.v.*，且分布函数在 0 点不连续），并与前两个例子作以比较，以加深对条件分布及存在 *p.d.f.*必要条件的理解。

注意：*p.d.f.* $f_X(x)$表征的是随机变量在某一点 x 处的概率特性，而建立在条件概率和

*p.d.f.*基础上的条件 *p.d.f.*表征的是在给定事件已发生的条件下，*r.v.*在某一点 x 处的条件概率特性。

在许多理论研究和实际应用中，通常关心 *r.v.*限制在某一范围内的概率分布，因此下面我们引入截尾分布函数的定义。

定义 5.2　对于一般的 *r.v.X*，称 X 在（$a<X<b$）条件下的条件概率分布为 X 的截尾分布函数，其中$-\infty\le a<b\le+\infty$。

易知，例 5.1 实际上是求一个服从正态分布的 *r.v.*的截尾分布函数。

例 5.2　某种型号的电子管寿命 X（以小时计）服从参数为 $\lambda=2$ 的指数分布，厂家规定：$X>4$ 的电子管为合格品，而 $X>M$ 的电子管为优质品，M 随顾客要求而定（$M>4$）现随机抽取一个电子管，已知它为合格品，求其是优质品的概率（此概率是 M 的函数）。

解：　本题也就是求符合指数分布的 *r.v.*的截尾分布函数。

由于 $X\sim E(\lambda)$，则 $f_X(x)=\lambda e^{-\lambda x}I_{(x\ge0)}$，由题意 $\lambda=2$

$$P(X\le x)=1-e^{-2x}，P(X>x)=e^{-2x}$$

当 $M>4$ 时，$P(X>M|X>4)=\dfrac{P(X>M)}{P(X>4)}=\dfrac{e^{-2M}}{e^{-2\times4}}=e^{-2(M-4)}$

可见，X 在 $X>4$ 的条件下仍服从指数分布。

在第三节中我们介绍了指数分布具有无记忆性，上例说明了同一性质。

条件分布函数和条件 *p.d.f.*在概率论中具有很重要的地位，因为在实际应用中，人们往往关心的是随机变量在某一特定范围内的概率特性。例如仅考虑男生的身高变化。这相对于全班同学的身高，就是有条件的，涉及到条件分布函数和条件 *p.d.f.*的相应问题。

此外，深入掌握一维 *r.v.*，对以后学习多维 *r.v.*及条件数学期望都是很有益处的。

§6 随机变量函数的分布

在实际问题中，我们往往会对某些随机变量的函数比对该随机变量更感兴趣。例如，在一些试验中，所关心的随机变量往往不能由直接测量得到，而它却是某个能直接测量的随机变量的函数。例如，我们能测量圆的直径 D，而关心的却是圆的面积 $S=\frac{1}{4}\pi D^2$。这里 *r.v.S* 是 *r.v.D* 的函数。

在这一节中，要讨论的问题是：已知 $X(\omega)$的概率分布，求它的函数 $Y(\omega)=g(X(\omega))$的概率分布，其中 $g(\cdot)$是已知的一元实值函数。然而，对任意 *r.v.*$X(\omega)$，是否对任意一个实函数 $g(\cdot)$，$g(X(\omega))$还是 *r.v.*呢？事实上，要求 $g(x)$为波雷尔（*Borel*）可测函数，即：

对 $\forall y \in R$，有 $\{x : g(x) \le y\} \in$ ✍。

容易验证，单调函数是波雷尔可测函数，逐段单调函数、连续函数等均是波雷尔可测函数。我们有以下命题：

命题：设 $X(\omega)$是概率空间(Ω, Φ, P)上的一个随机变量，$g(x)$是波雷尔可测函数，则 $Y(\omega) \mathrel{\widehat{=}} g(X(\omega))$ 也是概率空间(Ω, Φ, P)上的一个随机变量。

证明：要证 $\forall a \in R, \{\omega : g(X(\omega)) \le a\} \in$ ☞。记 $B = \{x : g(x) \le a\}$，由于 $g(x)$是波雷尔可测函数，故 $B \in$ ✍。下证 $\{\omega : g(X(\omega)) \le a\} = \{\omega : X(\omega) \in B\}$：

$$\forall \omega_1 \in \{\omega : g(X(\omega)) \le a\} \;\square\; g(X(\omega_1)) \le a \;\square\; X(\omega_1) \in B \;\square\; \forall \omega_1 \in \{\omega : X(\omega) \in B\}$$

故 $\{\omega : g(X(\omega)) \le a\} = \{\omega : X(\omega) \in B\}$；于是 $\{\omega : X(\omega) \in B\} \in$ ☞ □ $\{\omega : g(X(\omega)) \le a\} \in$ ☞
即 $Y(\omega) = g(X(\omega))$ 是概率空间(Ω, Φ, P)上的一个随机变量。

下面，我们讨论几个具体例子。

例 6.1 设 *d.r.v.*分布律如下，求 $Y=X^2+1$ 的分布律。

X	-2	-1	0	1	2
P	2/15	3/15	5/15	4/15	1/15

解：Y 所有可能取值为 1，2，5，故：

$$P\{Y=1\} = P\{X^2+1=1\} = P\{X=0\} = \tfrac{5}{15} = \tfrac{1}{3}$$

$$P\{Y=2\} = P\{X^2+1=2\} = P\{X=1\} + P\{X=-1\} = \tfrac{4}{15} + \tfrac{3}{15} = \tfrac{7}{15}$$

$$P\{Y=5\} = P\{X^2+1=5\} = P\{X=2\} + P\{X=-2\} = \tfrac{1}{15} + \tfrac{2}{15} = \tfrac{1}{5}$$

则 Y 的分布律为：

Y	1	2	5
P	1/3	7/15	1/5

例 6.2　设 $r.v.X \sim N(0,1)$ 求：$Y=X^2$ 的 *p.d.f.*。

解：设 Y 的分布函数 $F_Y(y)=P(Y \le y)$。由于 $Y=X^2 \geqslant 0$，故 $y<0$ 时，$F_Y(y)=0$；
当 $y \geqslant 0$ 时，

$$P\{Y \le y\} = P\{X^2 \le y\} = P\{-\sqrt{y} \le X \le \sqrt{y}\} = \int_{-\sqrt{y}}^{\sqrt{y}} f_X(x)dx$$

用 $f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ 代入上式，然后对 y 求导，即得 $Y=X^2$ 概率密度函数为：

$$f_Y(y) = \frac{y^{-1/2}}{\sqrt{2\pi}} e^{-\frac{y}{2}} \cdot I_{(y \ge 0)}$$

此时称 Y 服从自由度为 1 的卡方分布。（见第三章练习题 3.30）

例 6.3 设 $r.v.\ X \sim U(0,1)$，求 $Y=\Phi^{-1}(X)$的概率密度函数(其中 $\Phi^{-1}(x)$是标准正态分布函数$\Phi(x)$的反函数)。

解：先求 $Y=\Phi^{-1}(X)$的分布函数 $F_Y(y)=P(Y\le y)$，注意 $X\sim U(0,1)$,故有 $F_Y(y)=P(Y\le y)=P(\Phi^{-1}(X)\le y)=P(X\le\Phi(y))=\Phi(y)$ 知 $Y \sim N(0,1^2)$,得Y的$p.d.f$为$f_Y(y)=\varphi(y)=\dfrac{1}{\sqrt{2\pi}}e^{-y^2/2}$。

可知，由服从(0,1)上均匀分布的 $r.v.\ X$，经变换 $Y\equiv\Phi^{-1}(X)$得到$r.v.Y \sim N(0,1^2)$，这为产生服从正态分布的 $r.v.$提供了一个途径。

以上三例都是直接求其分布律或分布函数。这种解法可简称为“直接法”。它的关键在于事件的转化，即如何将 $r.v.Y$ 表示的事件转化为 $r.v.X$ 表示的事件。

例 6.4 设 $r.v.X \sim N(\mu,\sigma^2)$，试证明 X 的线性组合 $Y=aX+b\ (a\neq0)$也服从正态分布。

解： 先求 Y 的分布函数，即需将事件$\{\omega：Y\leqslant y\}$ 转化为 $r.v.X$ 表示的事件。

先设 $a>0$,

$$F_Y(y)=P\{Y\le y\}=P\{aX+b\le y\}=P\{X\le\frac{y-b}{a}\}=\int_{-\infty}^{\frac{y-b}{a}}\frac{1}{\sqrt{2\pi}o}e^{-\frac{(x-\mu)^2}{2o^2}}dx$$

对 $F_Y(y)$求导，得 $p.d.f.$

$$f_Y(y)=\frac{1}{a}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(\frac{y-b}{a}-\mu)^2}{2\sigma^2}}=\frac{1}{\sqrt{2\pi}a\sigma}e^{-\frac{(y-b-a\mu)^2}{2(a\sigma)^2}}\text{，}-\infty<y<+\infty$$

对 $a<0$:

$$P(Y\le y)=P(aX+b\le y)=P(X\ge\frac{y-b}{a})=1-\int_{-\infty}^{(y-b)/a}\frac{1}{\sqrt{2\pi}\sigma}e^{-(x-\mu)^2/2\sigma^2}dx,$$

对y求导得:

$$f_Y(y)=\frac{1}{\sqrt{2\pi}|a|\sigma}e^{-(y-b-a\mu)^2/2(a\mu)^2}$$

故 $Y=aX+b\sim N(a\mu+b，a^2\sigma^2)$。

可见 $r.v.Y$ 仍服从正态分布，其参数为 $a\mu+b$ 和 $a^2\sigma^2$。

结论：服从正态分布的随机变量的线性函数也服从正态分布。

还可以采用“微元法”求 $r.v.$函数的概率密度函数，在本章第三节，我们曾提到微元法的依据，即：当 $f_Y(y)$ 在 y 连续时，有：

$$f_Y(y)=\underset{h\to 0}{Lim}\frac{P(y<Y\le y+h)}{h}$$

这里也可以应用这种关系，对例 6.4 求解。

例 6.5 题设同上例。

解：用微元法。先设 a>0:

$\forall h>0, h$充分小，则

$$P(y<Y\le y+h)=P(y<aX+b\le y+b)=P(\frac{y\text{-}b}{a}<X\le\frac{y\text{-}b}{a}+\frac{h}{a})=f_X(\frac{y\text{-}b}{a})\cdot\frac{h}{a}+o(\frac{h}{a})$$

故：

$$f_Y(y)=\underset{h\to 0}{Lim}\frac{P(y<Y\le y+h)}{h}=\underset{h\to 0}{Lim}\frac{f_X(\frac{y\text{-}b}{a})\cdot\frac{h}{a}}{h}=f_X(\frac{y\text{-}b}{a})\cdot\frac{1}{a}$$

$$=\frac{1}{\sqrt{2\pi}\sigma\cdot a}\exp\left\{-\frac{[y-(a\mu+b)]^2}{2(a\sigma)^2}\right\}$$

当 a<0 时,

$$P(y<Y\le y+h)=P(\frac{y-b}{a}>X\ge\frac{y-b}{a}+\frac{h}{a})=f_X(\frac{y-b}{a})(-\frac{h}{a})+O(\frac{h}{a})$$

$$\Rightarrow f_Y(y)=\frac{1}{\sqrt{2\pi}|a|\sigma}e^{-\frac{[y-(a\mu+b)]^2}{2(a\sigma)^2}}$$

得 $Y=aX+b\sim N(a\mu+b,(a\sigma)^2)$。

例 6.6 设电压 $V=A\sin\theta$，其中 A 是已知的正常数，相角 θ 是一 $r.v.$。$\theta\sim U(0,\pi)$，试求电压 V 的概率密度。

解：用微元法。

$\forall v\in(0,\pi)$ 取 h>0 充分小，使 $v+h\in(0,\pi)$

$P\{v<V\le v+h\}=P\{v<A\sin\theta\le v+h\}$

$$=P\{\arcsin\tfrac{v}{A}<\theta\le\arcsin\tfrac{v+h}{A}\}+P\{\pi-\arcsin\tfrac{v+h}{A}\le\theta<\pi-\arcsin\tfrac{v}{A}\}$$

$$=\tfrac{1}{\pi}(\arcsin\tfrac{v+h}{A}-\arcsin\tfrac{v}{A})+\tfrac{1}{\pi}(\arcsin\tfrac{v+h}{A}-\arcsin\tfrac{v}{A})$$

由台劳展开： $$\arcsin\tfrac{v+h}{A}=\arcsin\tfrac{v}{A}+\frac{1}{\sqrt{1-\frac{v^2}{A^2}}}\cdot\frac{1}{A}\cdot h+\circ(h)$$

因此有： $$P(v<V\le v+h)=\frac{2}{\pi}\cdot\frac{1}{\sqrt{A^2-v^2}}\cdot h+\circ(h)$$

所以 $$f_V(v)=\underset{h\to 0}{Lim}\frac{P(v<V\le v+h)}{h}=\frac{2}{\pi\sqrt{A^2-v^2}}$$

因此
$$f_V(v)=\begin{cases}\dfrac{2}{\pi\sqrt{A^2-v^2}} & -A<v<A\\ 0 & 其它\end{cases}$$

例 6.7 设 $X\sim N(2, 1^2)$，$B=\{\omega: X\geq 0\}$。求 X 在 B 条件下的条件概率密度函数 $f_{X|B}(x)$。

解：当 $x\leq 0$ 时，$P(X\leq x, X\geq 0)=0$，故 $x\leq 0$ 时，有 $f_{X|B}(x)=0$；故只需考虑 $x>0$ 的情形。对 $x>0$，取 $h>0$ 充分小，则：

$$\begin{aligned}P(x<X\leq x+h\mid X\geq 0)&=P(x<X\leq x+h)/P(X\geq 0)=(1-\phi(-2))^{-1}P(x<X\leq x+h)\\&=\phi(2)^{-1}\int_x^{x+h}(2\pi)^{-1/2}\exp(-(u-2)^2/2)du\\&=\phi(2)^{-1}[(2\pi)^{-1/2}\exp(-(x-2)^2/2)h+o(h)]\end{aligned}$$

$$\therefore f_{X|X\geq 0}(x)=\lim_{h\to 0}\frac{P(x<X\leq x+h)}{h}=(0.9772)^{-1}(2\pi)^{-1/2}\exp(-(x-2)^2/2)$$

故：$f_{X|X\geq 0}(x)=(0.9772)^{-1}(2\pi)^{-1/2}\exp(-(x-2)^2/2)I_{(X>0)}$。

例 **6.8**（柯西分布）设 $r.v.\theta\sim U(-\frac{\pi}{2},\frac{\pi}{2})$，令 $Z=r\,tg\theta$（其中 $r>0$ 为常数），试求 Z 的 $p.d.f.$。

解：注意到 $Z=rtg\theta$ 是单调增函数，因此 $\forall x\in(-\infty,\infty)$，

$$\begin{aligned}P(Z\leq x)&=P(rtg\theta\leq x)=P(tg\theta\leq\frac{x}{r})=P(-\frac{\pi}{2}\leq\theta\leq arctg\frac{x}{r})=\frac{1}{\pi}(arctg\frac{x}{r}+\frac{\pi}{2})\\&=\frac{1}{\pi}\cdot arctg\frac{x}{r}+\frac{1}{2}\end{aligned}$$

对上式求导可知：Z 的 $p.d.f.$ 为 $f(x)=(\frac{1}{\pi}arctg\frac{x}{r}+\frac{1}{2})'_x=\frac{1}{\pi}\cdot\frac{r}{r^2+x^2}$　。

综上，对于求 $r.v.$函数的分布函数与 $p.d.f.$问题，我们可以用“直接法”或“微元法”解决。读者要体会二者的关键及实质。

练习题

2.1 一批零件中有九个合格品三个不合格品。安装配件时，从这批零件中任取一个，如果每次取出的不合格品不再放回，而再取一个零件，直到取得合格品为止。求在取得合格品以前以取出不合格品数的概率分布。

2.2 先掷一匀称的骰子，然后抛个数与骰子掷出的点数相同的硬币。

（1）求国徽朝上个数的分布列；

（2）若得到 3 个国徽朝上，问骰子掷出 n 点的概率多大？

2.3 掷一颗匀称的骰子 n 次，设 M 与 m 分别表示所得点子的最大值与最小值，求：$P(m=2, M=5)$。

2.4 由某机器生产的螺栓的长度（cm）服从参数为 $\mu=10.05$，$\sigma=0.06$ 的正态分布，规定长度在范围 10.05 ± 0.12 内为合格品。求一螺栓为不合格品的概率。

2.5 已知 $lnX\sim N(1,2^2)$，求 $P(\frac{1}{2}<X<2)$。

2.6 点随机地落在中心在原点、半径为 R 的圆周上, 并且对弧长是匀称地分布的, 求落点的横坐标的概率密度。

2.7 设 X 的分布函数 $F(x)$是连续函数, 则 $Y=F(X)$服从 $U(0,1)$。

2.8 抛掷一枚分币，直到出现“国徽朝上”为止，求抛掷次数的概率分布。

2.9 从一副扑克牌中发出 5 张，求其中黑桃张数的概率分布。

2.10 甲、乙二人投篮，投中的概率分别为 0.6，0.7，今各投 2 次，求：

（1）二人投中次数相等的概率；（2）甲比乙投中次数多的概率。

2.11 设 X 服从泊松分布，已知 $P(X=1)=P(X=2)$，求 $P(X=4)$。

2.12 设 X 服从泊松分布（参数为 λ），问：k 取何值时，$P(X=k)$ 为最大？

2.13 设 $X\sim B(n, p)$。问：k 取何值时，$P(X=k)$ 为最大？

2.14 在一次核反应中某个粒子可能分裂为 2 或 3 个粒子或不分裂，这三种可能性相应的概率分别为 p_2, p_3和p_1，新粒子的分裂性态是相同的（与老粒子也相同），并且分裂结果彼此独立。求两次反应以后粒子总数的分布。

2.15 设随机变量 X 的概率密度为 $f(x)=CxI_{(0\le x\le 1)}$，求：（1）常数 C；（2）X 落在区间（0.3，0.7）内的概率。

2.16 设 $X\sim N(1,0.2^2)$，求：（1）$P(X>0)$；(2)$P(0.2<X<1.8)$。

2.17 设 $X\sim U(-\frac{\pi}{2},\frac{\pi}{2})$，证明：$Y=\tan X$ 的概率密度函数为 $f_Y(y)=(\pi(1+y^2))^{-1}$，$-\infty<y<+\infty$，称 Y 是服从 *Cauchy* 分布的随机变量。

2.18 设随机变量 X 的概率密度 $f(x)$ 为：

（1）$f(x)=\begin{cases}\frac{1}{\pi}\sqrt{1-x^2}, & -1\le x\le 1；\\ 0, & 其他；\end{cases}$ （2）$f(x)=\begin{cases}x, & 0\le x<1；\\ 2-x & 1\le x\le 2；\\ 0, & 其他。\end{cases}$

分别求出 X 的分布函数 $F(x)$，并作出（2）中的 $f(x)$ 与 $F(x)$ 的图形。

2.19 设 *r.v.X* 服从参数为 p $(0<p<1)$ 的几何分布，而 $X\sim Ge(p),\forall l\in N_0$，求在 $X\geq l$ 条件下的条件分布律。

2.20 设 *r.v.X* 只取非负整数值，若对任意 $l,k\in N_0$，有 $P(X\ge l+k|X\ge k)=P(X\ge l)$，试证：$X$ 服从几何分布。

2.21 设连续型 *r.v.X* 的分布函数 $F(x)\equiv P(X\le x)$，试证：*r.v.* $Y\equiv F(X)$ 服从均匀分布 $U(0,1)$。

2.22 称 *r.v.X* 服从威布尔分布(*Weibull distribution*)，若它的 *p.d.f* 为：

$f(x)=\alpha mx^{m-1}\exp(-\alpha x^m)$ $x\ge 0,\alpha>0,m>0$（此分布在可靠性理论中常用）。

试求 $Y=X^m$ 的 *p.d.f.*。

2.23 $F(x)$ 是一分布函数，定义：$\forall y\in(0,1),F^{-1}(y)\equiv\sup\{x:F(x)<y\}$。已知 $\forall x\in R$，$y\in(0,1)$ 满足 $\left(F^{-1}(y)\le x\right)$（其充要条件为 $(y\le F(x))$，对此事实有兴趣的读者可作为练习加以证明。）试证：若 $X\sim U(0,1)$，则 $Y\equiv F^{-1}(x)$ 的分布函数为 $F(x)$。(此结果在随机模拟中有重要应用。)

2.24 设 *r.v.X*～$U(0,1)$，令 $Y=-\lambda^{-1}\ln X$，其中 $\lambda>0$ 为常数，求 Y 的 *p.d.f.*。

2.25 设 *r.v.X*～ $N(0,1^2)$，求 $Y=X^2$ 及 $Z=\sqrt{Y}$ 的 *p.d.f.*。

2.26 设 *r.v.X*～$Ex(\lambda)$，给定 $l>0$，求 X 在 $(X\geq l)$ 条件下的条件概率密度函数。

2.27 设 *r.v.Y*～$U[0,1]$，　令 $Y_n=\sum_{k=0}^{2^n-1}\left(k/2^n\right)\cdot I_{(k/2^n\le Y<k+1/2^n)},n\ge 0$。求：

(1)分别求 Y_1,Y_2 的分布律;

(2) Y_n 的分布律;

(3) Y_2 在给定 $Y_1=1/2$ 下的条件分布律。

2.28 设 *r.v.*X～$B(n,p)$，试用示性函数的线性组合来表示 *r.v.*X。

2.29 设 *r.v.*Y～$Po(\lambda)$，试用示性函数的线性组合来表示 *r.v.*Y。

2.30 设一批电子元件某指标 X～$\mathrm{N}(50,1^2)$（设批量为∞），按质量要求：当$|X-50|\le 2$为合格品，否则为次品。求将产品中不合格元件全部剔除后，剩下元件指标的 *p.d.f.*。

2.31 某批零件强度 X～$N(48,2^2)$（设批量为∞），若 $X\ge 46.32$，则此零件为合格品，否则为次品。制定抽验方案如下：第一次任取 3 个，若 3 个都合格，则接收该批零件；若至少有 2 个次品，则拒收该批；否则再抽第二次，也是 3 个。重复上述步骤直至作出接收或拒绝决定为止。(1)求第一次抽验 3 个而接收该批零件的概率及第一次抽验而作出决定的概率；(2)求通过抽验接收该批产品的概率；(3)记 N 为直至作出决定的抽取次数,求 N 的分布律。

2.32 设某电子元件在工作时，其两端电压 V～$N(220,400)$。当 $V\in[200,240]$时，其失效概率为 0.05；当 $V<200$ 时，其失效概率为 0.10；当 $V>240$ 时，其失效概率为 0.50，求：(1)此元件的失效概率；(2)当元件失效时，电压超过 240 的条件概率。

第三章　多维随机变量及其分布

上一章，我们研究了单一的随机变量，引出了刻划它的概率特性的若干概念与性质，从而使描述随机现象变得清晰而简洁。但是在许多随机试验中，每次试验的结果需用几个量来表达。例如，考察一次射击的弹着点，需用两个量(纵、横坐标)表示。又如，考察一个“服务台”的情况，需同时考察每个“顾客”到达时刻、等待时间和被服务的时间，等等。这样对于每个样本点，试验结果必须用一个随机向量$(X_1(\omega),X_2(\omega),\ldots,X_n(\omega))\in\boldsymbol{R}^n$来表示。对于这样的随机向量，不仅要研究各自的特性，又要研究它们之间的各种关系，后者有时尤为重要，因此需将它们作为整体来研究。

定义： 称概率空间(Ω,Φ,P)上n个随机变量$X_1,X_2,\ldots,X_n$的整体$(X_1,X_2,\ldots,X_n)$为 **n 维随机向量**，也称 **n 维随机变量。**

上面的例子中炮弹的弹着点就是一个二维随机向量，而考察前n个“顾客”相继到达时刻，则是n维随机向量。

在本章中，我们将对多维随机向量(以二维为主)进行研究，给出它的一些研究方法和性质。

Comment:

§1　离散型随机变量及其分布

1　联合分布律

定义 1.1　给定(Ω,Φ,P),

（1）若$X=(X_1,X_2,\ldots,X_n)$每一个分量都是离散型随机向量，则称 $\boldsymbol{X}$ 为 **n 维离散型随机变量**。

（2）若二维离散性随机变量(X,Y)可能取值为$(x_i,y_j),i,j\in N$；则称

$$\{p_{ij}=P(X=x_i,Y=y_j),i,j\in N\}$$

为(X,Y)的**联合分布律**，简称**分布律**。

由定义可知二维离散型随机变量(X,Y)的联合分布律具有以下性质：　$p_{ij}\geq 0$；$\sum_i\sum_j p_{ij}=1$。并且具有上述性质的任何数集$\{p_{ij};\ i,j=1,2,\ldots\}$均可作为某二维离散型随机变量的分布律。

2　边缘分布与条件分布

为简单计，只讨论二维 $d.r.v.$的情形，多维情形的定义和结论完全类似。

定义 1.2 设(X,Y)的联合分布为 $P\left((X,Y)=(x_i,y_j)\right)=p_{ij},i,j\in \mathrm{N}$

（1）称 $P(X=x_i),i\in \mathrm{N}$ 为 X 的**边缘分布**，称 $P(Y=y_j),j\in \mathrm{N}$. 为 Y 的**边缘分布**。

（2）若 $P(Y=y_j)>0,j\in N$，称 $P(X=x_i\mid Y=y_j),i\in N$，为 X 关于 $(Y=y_j)$ 的**条件分布**。

易知：$P(X=x_i)=\sum_{j=1}^{\infty}p_{ij}$，$P(Y=y_j)=\sum_{i=1}^{\infty}p_{ij}$，$P(X=x_i\mid Y=y_j)=p_{ij}\big/\sum_i p_{ij}$ 。

例 1.1 袋中装有 2 只白球，3 只黑球，分别进行有放回及无放回的抽球，设：

$$X=\begin{cases}1 & \text{第一次抽出白球}\\ 0 & \text{第一次抽出黑球}\end{cases}\qquad \mathrm{Y}=\begin{cases}1 & \text{第二次抽出白球}\\ 0 & \text{第二次抽出黑球}\end{cases}$$

分别求(X,Y)的联合概率分布及边缘分布律。

解： (1). 作放回抽取，第一次抽取和第二次抽取是独立的

$$\therefore P\{X=x_i,Y=y_i\}=P\{X=x_i\}P\{Y=y_i\},i=1,2;j=1,2$$

列表如下：

$Y\setminus X$	0	1	$P\{Y=y_i\}$
0	$\frac{3}{5}\cdot\frac{3}{5}$	$\frac{2}{5}\cdot\frac{3}{5}$	$\frac{3}{5}$
1	$\frac{3}{5}\cdot\frac{2}{5}$	$\frac{2}{5}\cdot\frac{2}{5}$	$\frac{2}{5}$
$P\{X=x_i\}$	$\frac{3}{5}$	$\frac{2}{5}$	/

(2). 做无放回抽取，列表如下：

$Y\setminus X$	0	1	$P\{Y=y_i\}$
0	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{2}{5}\cdot\frac{3}{4}$	$\frac{3}{5}$
1	$\frac{3}{5}\cdot\frac{2}{4}$	$\frac{2}{5}\cdot\frac{1}{4}$	$\frac{2}{5}$
$P\{X=x_i\}$	$\frac{3}{5}$	$\frac{2}{5}$	/

我们看到，放回和不放回两种抽样方式，尽管联合概率分布不同，但边缘分布却是一致的。

3. 和、差、积、商、极大、极小

二个离散型随机变量（$X(w),Y(w)$）那么它们的和、差、积、商，极大值与极小值是否是随机变量？回答是肯定的。

命题 1.1 设$(X.Y)$是定义在(Ω, Φ, P)上的二维离散型随机变量，

则(1)$X+Y$，(2)$X-Y$，(3)XY，(4)$X/Y(Y\neq 0)$，(5) $X\vee Y$，(6) $X\wedge Y$ 均是离散型随机变量：

证明： 设 $d.r.v.(X,Y)$的分布律为：

$$P(X=x_i, Y=y_j)=p_{ij}\geq 0,(i,j\in N)$$

则$\forall a\in R\ (x_i, y_j), i, j\in N$有：$(X=x_i)\in\Phi$，$(Y\leq a-x_i)\in\Phi$，$\Phi$为σ—域，故：

$$(X+Y\leq a)=\bigcup_i (X=x_i, Y\leq a-x_i)\in\Phi,$$

类似地，

$$(X-Y\leq a)=\bigcup_j (Y=y_j, X\leq a+y_j)\in\Phi,$$

$$(X\cdot Y\leq a)=\bigcup_{i:x_i>0}(X=x_i, Y\leq a/x_i)\cup\bigcup_{i:x_i<0}(X=x_i, Y\geq\frac{a}{x_i})\in\Phi,$$

当 $r.v.\ Y\neq 0$ 时 $\{(X/Y)\leq a\}=\bigcup_{j:y_j>0}\{Y=y_j, X\leq ay_j\}\cup\bigcup_{j:y_j<0}\{Y=y_j, X\geq ay_j\}\in\Phi$，

$\{(X\vee Y)\leq a\}=(X\leq a, Y\leq a)\in\Phi$，$\{(X\wedge Y)>a\}=(X>a, Y>a)\in\Phi$。

例 1.2 设离散型 $r.v.(X,Y)$的联合分布律如下：

(X,Y)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)	(2,3)
P_{ij}	$1/18$	$5/18$	$3/18$	$3/18$	$4/18$	$2/18$

求(1)$X+Y$ 的分布律，(2)$X\vee Y$ 的分布律

解(1)$X+Y$ 的可能取值为 2~5,

$P(X+Y=2)=P(X=1, Y=1)=1/18,$

$P(X+Y=3)=P(X=1, Y=2)+P(X=2, Y=1)=8/18$

$P(X+Y=4)=P(X=1, Y=3)+P(X=2, Y=2)=7/18$

$P(X+Y=5)=P(X=2, Y=3)=2/18$故

$X+Y$	2	3	4	5
P_i	$1/18$	$8/18$	$7/18$	$2/18$

(2) $X\vee Y$ 可能取值为 1,2,3，

$$P((X\vee Y)=1)=1/18,\quad P((X\vee Y)=2)=8/18,\quad P((X\vee Y)=3)=5/18\ .$$

3 离散型 *r.v.*的独立性

这里只讨论三维情形，$n \geq 4$ 情形完全相类似。

定义 1.3 设(X,Y,Z)是三维离散型随机变量，联合分布为

$P(X = x_i, Y = y_j, Z = z_k) = p_{ijk}$，$i, j, k \in N$；若 $\forall i, j, k \in N$ 均有

$$P(X = x_i, Y = y_j, Z = z_k) = P(X = x_i)P(Y = y_j)P(Z = z_k)$$

则称 *d.r.v.*(X,Y,Z)相互独立。

以下略举它的若干性质。

命题 1.1 设 *d.r.v.*X,Y,Z 相互独立，则

（1）$\forall i, j, k \in N$，$P(X = x_i \mid Y = y_j) = P(X = x_i)$，$P(X = x_i \mid Z = z_k) = P(X = x_i)$，

$P(Y = y_j \mid Z = z_k) = P(Y = y_j)$，$P(X = x_i \mid Y = y_j, Z = z_k) = P(X = x_i)$

（2）$X \pm Y$与Z独立，XY、$\dfrac{X}{Y}$分别与Z独立，$X^2 + Y^2$与Z^2相互独立。

证略。留给读者作为练习。

以上只列举了 X,Y,Z 相互独立的部分性质，读者有兴趣的话，可进一步列出其它性质。

§2 连续型随机变量及其概率密度

1 联合概率密度函数

定义 2.1 给定(Ω,Φ,P)，若$\exists f(x_1,..,x_n) \geq 0$，对$\forall B \in$ ✍^{n}，有

$$P(\omega : X(\omega) \in B) = \int \cdots \int_B f(x_1,...,x_n)dx_1...dx_n$$

则称 $X(\omega)$ 为 **n 维连续型随机变量**，$f(x_1,...,x_n)$为 $X(\omega)$的**联合概率密度函数**，简记为 *j.p.d.f*。

这里着重讨论二维情形，n 维（$n \geq 3$）情形完全类似。

二维 *r.v.s* (X,Y)的 *j.p.d.f* $f(x,y)$具有以下性质：

(1) 非负性：$f(x,y) \geq 0$；

(2) 规一性：$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy = 1$；

(3) 设 $G\in \mathscr{B}^2$ 是 xOy 平面上的一个区域，点(X,Y)落在 G 内的概率为：

$$P\{(X,Y)\in G\}=\iint_G f(x,y)dxdy$$

几何上的值等于以 G 为底，以 $z=f(x,y)$为顶面的柱体体积。其中$\mathscr{B}^2$为二维 *Borel* $\sigma-$域。

(4) 若 $f(x,y)$在(x,y)连续，则：

$$f(x,y)=\lim_{\substack{h_1\to 0\\ h_2\to 0}}\frac{P(x<X\le x+h_1,y<Y\le y+h_2)}{h_1h_2} \qquad (2.1)$$

利用（2.1）式，求(X,Y)的 *j.p.d.f.*。此法称为微元法，详见本章§6、§7 及第六章等。

2　边缘密度函数

对于二维随机变量(X,Y)，其分量 X(或 Y)的密度函数称为**边缘密度函数**。

我们有：$f_X(x)=\int_{-\infty}^{\infty}f(x,y)dy$ 。

证明：因 $\forall B\in \mathscr{B}$，

$$P(X\in B)=\iint_{x\in B}f(x,y)dxdy=\int_{x\in B}(\int_{-\infty}^{+\infty}f(x,y)dy)dx,$$

由上一章$p.d.f.$定义可知$\int_{-\infty}^{\infty}f(x,y)dy$为$X$的密度函数。

例 2.1　二维随机变量（X,Y）在 $G_R=\{(x,y):x^2+y^2\le R^2\}$ 上服从二维均匀分布，

即 $f(x,y)=\dfrac{1}{\pi R^2}I_{(x^2+y^2\le R^2)}$，求$f_X(x),f_Y(y)$。

解：由边缘分布的性质：

当$|x|>R$时　$f_X(x)=0$；

当$|x|\le R$时　$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\int_{-\sqrt{R^2-x^2}}^{\sqrt{R^2-x^2}}\dfrac{1}{\pi R^2}dy=\dfrac{2\sqrt{R^2-x^2}}{\pi R^2}$。

即：　$f_X(x)=\dfrac{2\sqrt{R^2-x^2}}{\pi R^2}I_{(|x|\le R)}$。

同理可求　$f_Y(y)=\dfrac{2\sqrt{R^2-y^2}}{\pi R^2}I_{(|y|\le R)}$。

由此可知，联合分布为均匀分布，边缘分布未必也是均匀分布。

3 二维正态随机变量

定义 2.2 若二维随机变量(X_1, X_2)的 $j.p.d.f$ 为：

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\} \tag{2.2}$$

其中$\mu_1,\mu_2,\sigma_1,\sigma_2,\rho$都是常数，且$\sigma_1>0,\sigma_2>0,-1\le\rho\le1$ 。则称 $(X_1,\ X_2)$ 是参数为 $(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$ 的二维正态随机变量。简记为：

$$(X_1, X_2)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)。$$

命题 2.1 设(X_1, X_2)为如上二维正态，则 $X_k\sim N(\mu_k,{\sigma_k}^2),1\le k\le 2$ 。

解：

$$f(x_1,x_2)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{\rho^2(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]+\frac{-(x_1-\mu_1)^2}{2\sigma_1^2}\right\}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{\frac{-(x_1-\mu_1)^2}{2\sigma_1^2}\right\}\cdot\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\cdot\exp\left\{\frac{-1}{2\sigma_2^2(1-\rho^2)}\left[(x_2-\mu_2)-\rho\frac{\sigma_2}{\sigma_1}(x_1-\mu_1)\right]^2\right\}$$

由边缘 $p.d.f.$与联合 $p.d.f.$的关系有 $f_{X_1}(x_1)=\int_{-\infty}^{+\infty}f(x_1,x_2)dx_2$

令$t=(1-\rho^2)^{-\frac{1}{2}}\sigma_2^{-1}[(x_2-\mu_2)-\rho\sigma_2\sigma_1^{-1}(x_1-\mu_1)]$，则有：

$$f_{X_1}(x_1)=\frac{1}{\sqrt{2\pi}\sigma_1}e^{\frac{-(x_1-\mu_1)^2}{2\sigma_1^2}}\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\frac{1}{\sqrt{2\pi}\sigma_1}e^{\frac{-(x_1-\mu_1)^2}{2\sigma_1^2}} \tag{2.3}$$

同理得：$f_{X_2}(x_2)=\frac{1}{\sqrt{2\pi}\sigma_2}e^{\frac{-(x_2-\mu_2)^2}{2\sigma_2^2}}$ ，而 X_k 的边缘分布为一元正态分布（$1\le k\le 2$）。

由这个命题知，二维正态分布的两个边缘分布都是一维正态分布，且都不依赖于参数 ρ。这说明单由关于 X_1 和 X_2 的边缘分布，一般是不能确定 X_1 和 X_2 的联合 $p.d.f.$。说明联合 $p.d.f.$所反映的信息，不仅反映各自的特征信息，还包含着二者之间某种关系的信息。

§3 联合分布函数

1 联合分布函数

定义 3.1 设(X,Y)为二维随机向量，$\forall(x,y)\in R^2$ 称 $F(x,y)=P(\omega:X\leqslant x,Y\leqslant y)$ 为二维随机向量(X,Y)的**联合概率分布函数**，或称为(X,Y)的联合**分布函数**。

对于二维 *r.v.*的联合分布，可给出其几何解释，若将二维 *r.v.*(X,Y)视为一个平面上的"随机点"的坐标，则 $F(x,y)$的函数值就是随机点(X,Y)落在如图 3-1 所示的阴影部分的概率。该概率是点(x,y)的二元函数。

基于上述分析,(X,Y)落于矩形区域(如图 3-2 所示)$[x_1<x\leqslant x_2,y_1<y\leqslant y_2]$的概率为

$$P(\omega: x_1<X\leqslant x_2,y_1<Y\leqslant y_2)=F(x_2,y_2)+F(x_1,y_1)-F(x_1,y_2)-F(x_2,y_1)$$

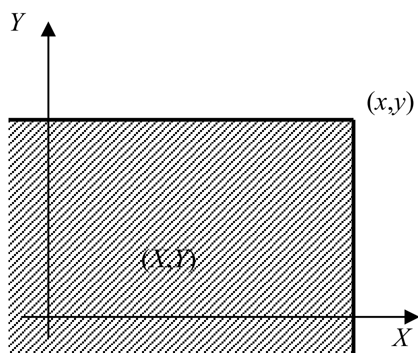


图 3-1

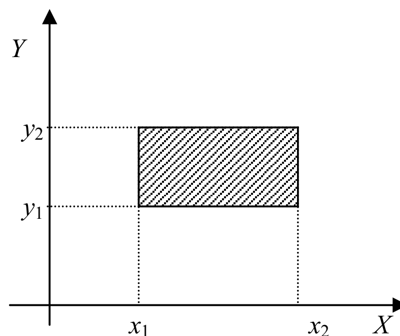


图 3-2

分布函数 $F(x,y)$具有以下的基本性质：

(1) $F(x,y)$是变量 x 和 y 的单调不减函数，即

$$\forall y\in R，若x_2>x_1,则F(x_2,y)\geq F(x_1,y)；\forall x\in R，若y_2>y_1,则F(x,y_2)\geq F(x,y_1)$$

其图形注释如下：

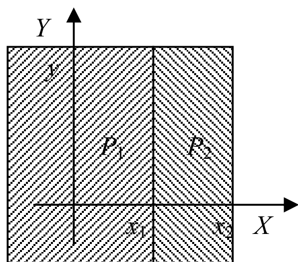


图 3-3 *a*

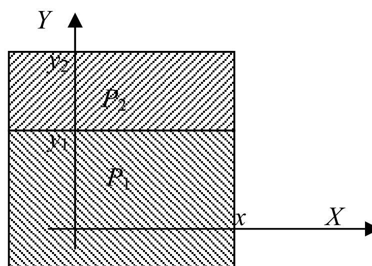


图 3-3 *b*

证明：

$$F(x_2,y)=P(X\le x_2,Y\le y)=P\{(X\le x_1,Y\le y)\bigcup(x_1\le X\le x_2,Y\le y)\}$$

$$=P(X\le x_1,Y\le y)+P(x_1<X\le x_2,Y\le y)\ge F(x_1,y)$$

（因为 $p=p_1+p_2$，且 $p_2\ge 0$，所以 $p\ge p_1$）

(2) $0\le F(x,y)\le 1$, $\lim\limits_{x\to-\infty}F(x,y)=0$；$\lim\limits_{y\to-\infty}F(x,y)=0$；$\lim\limits_{\substack{x\to+\infty\\y\to+\infty}}F(x,y)=1$

(3) $\forall x_1\le x_2, y_1\le y_2$；有 $F(x_2,y_2)-F(x_2,y_1)-F(x_1,y_2)+F(x_1,y_1)\ge 0$

证明：

左边 $=P(X\le x_2,Y\le y_2)-P(X\le x_2,Y\le y_1)-P(X\le x_1,Y\le y_2)+P(X\le x_1,Y\le y_1)$
$=P(x_1<X\le x_2,y_1<Y\le y_2)$
$=(X,Y)$落在图3－4所示的阴影部分的概率 ≥ 0

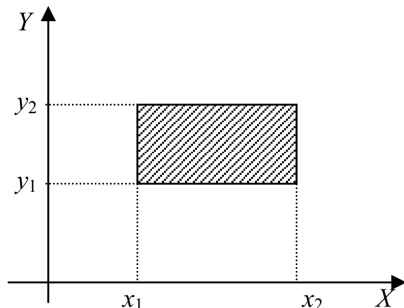


图 3-4

（4） 对于$\forall x\in R$，$F(x,y)$关于 y 右连续；对于$\forall y\in R$，$F(x,y)$关于 x 右连续。
证明：类似于一元分布函数右连续的证明，略。

2 边缘分布函数

定义 3.2 设二维随机变量（X,Y）的联合分布函数为 $F(x,y)$，而 X 和 Y 都是随机变量，相应的分布函数记为 $F_X(x),F_Y(y)$, 依次称为 X和 Y的**边缘分布函数**。简称边缘分布。

易知：

$$F_X(x)=F(x,+\infty)；F_Y(y)=F(+\infty,y) \qquad (3.1)$$

3 分布函数与密度函数的关系

若(X,Y)具有 j.p.d.f. f (x,y)，则 $\forall(x,y)\in R^2$，其联合分布和 X 的边缘分布为

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(u,v)dudv\text{，}\quad F_X(x)=\int_{-\infty}^{x}(\int_{-\infty}^{+\infty}f(u,v)dv)du \tag{3.2}$$

(3.2)式给出联合分布、边缘分布与密度函数的关系。可应用于已知密度函数，求其联合与边缘分布。反之，若密度函数未知，如何去求呢？以下几个命题为此提供略有不同的几种途径。

命题 3.1 设(X,Y)的联合分布函数为 F(x,y)，且除去可列点(x,y)外，若 $\dfrac{\partial^2 F}{\partial y\partial x}$ 在 $(x,y)\in R^2$ 上存在且连续，则(X,Y)的联合密度函数存在，且：

$$f(x,y)=\frac{\partial^2 F}{\partial y\partial x} \tag{3.3}$$

证明：类似于一维情形，从略。

命题 3.1 的条件太强，可稍减弱为：

命题 3.2 设(X,Y)的联合分布 F(x,y)在 R^2 上连续，且除去可列点(x,y)外（或除去一零测集（二维）外），$\dfrac{\partial^2 F}{\partial y\partial x}$ 存在且连续，则(X,Y)的 j.p.d.f.存在，且它可取为：

$$f(x,y)=\frac{\partial^2 F}{\partial y\partial x}I_A(u,v) \tag{3.4}$$

其中$A=\left\{(x,y):\dfrac{\partial^2 F}{\partial y\partial x}\text{存在且连续}\right\}$。

证明：类似于一维情形，从略。

(3.4)式意味着在 $\dfrac{\partial^2 F}{\partial y\partial x}$ 不存在的点上，即当 $(x,y)\notin A$ 时，取 f(x,y)=0。

有时(X,Y)的联合分布事先也未知，可用下法求其密度函数。

命题 3.2

设$r.v.(X,Y), A\subset R^2, A$是一个可列点集（或0测集(面积为0)),若$\forall(x,y)\in R^2$，
$P(X=x,Y=y)=0$,且$\forall(x,y)\in R^2-A$,极限

$$f(x,y)=\lim_{h_1,h_2\to 0}\frac{P(x<X\le x+h_1,y<Y\le y+h_2)}{h_1h_2} \tag{3.5}$$

存在连续，则(X,Y)的 j.p.d.f.存在且它可取为 $f(x,y)I_{R^2-A}(x,y)$。

证明：从略。

应用（3.5）式求 $f(x,y)$，称为微元法。

微元法可直观表述如下：为求 $f(x,y)$，先求 $P(x<X<x+h_1, y<Y<y+h_2)$。其次将上述概率分为二部分之和：一部分是与 $h_1\ h_2$成比例的主部，另一部分是 $o(h_1\ h_2)$，即是 $h_1\ h_2$的高阶无穷小，最后，第一项中 $h_1\ h_2$的系数即可作为(X, Y)的联合 $p.\ d.\ f.$。

关于微元法，在本章§6 中讨论 *r.v.*函数的分布及第六章中可以看到微小元法的重要作用。

4. 随机变量序列的极限表示

本小节讨论随机变量序列的上确界（下确界）上极限，下极限是否仍为 *r.v.*的问题，有如下：

命题 3.3 设 $\{X_n, n\ge 1\}$ 是定义在(Ω, Φ, P)上的 *r.v.*序列，则

$$\sup_n X_n(\omega)=\vee_n X_n(\omega),\quad \inf_n X_n(\omega)=\wedge_n X_n(\omega),$$

$$\lim_{n\to\infty}\sup X_n(\omega)\hat{=}\inf_{n\ge1}\sup_{k\ge n}X_k(\omega)=\wedge_{n\ge1}(\vee_{k\ge n}X_k(\omega))$$

及

$$\lim_{n\to\infty}\inf X_n(\omega)\hat{=}\sup_{n\ge1}\inf_{k\ge n}X_n(\omega)=\vee_{n\ge1}(\wedge_{k\ge n}X_k(\omega))$$

均为 *r.v*。

证明： $\forall a\in R$,因

$$\left\{\sup_n X_n\ (\omega)\le a\right\}=\bigcap_n(X_n(\omega)\le a)\in\Phi$$

$\left\{\inf_n X_n(\omega)>a\right\}=\bigcap_n(X_n(\omega)>a)\in\Phi$，故：$\left\{\inf_n X_n(\omega)\le a\right\}=\left\{\inf_n X_n(\omega)>a\right\}^C\in\Phi$

从而：$\lim_{n\to\infty}\sup X_n(\omega)=\inf_{n\ge1}\left(\sup_{k\ge n}X_k(\omega)\right)$ 与 $\lim_{n\to\infty}\inf X_n(\omega)\ =\sup_{n\ge1}(\inf_{k\ge n}X_k(\omega)\ )$ 均为随机变量，可知结论成立

命题 3.4$(X_n,n\geq 1)$同命题 3.1 定义。

（1）　记 $A=\left\{\omega:\lim_{n\to\infty}\inf X_n(\omega)=\lim_{n\to\infty}\sup X_n(\omega)\right\}$，则 $A\in\Phi$；

（2）　若 $\forall\omega\in\Omega$，有 $\lim_{n\to\infty}\inf X_n(\omega)=\lim_{n\to\infty}\sup X_n(\omega)$，则：

$$X(\omega)\hat{=}\lim_{n\to\infty}X_n(\omega)\hat{=}\lim_{n\to\infty}\sup X_n(\omega)$$

是 $r.v.$。

证明：(1) 因 $\overline{A}=\left\{\omega:\lim_{n\to\infty}\inf X_n(\omega)<\lim_{n\to\infty}\sup X_n(\omega)\right\}$

$$=\bigcup_k\left\{\omega:\liminf X_n\leq\frac{1}{k}<\limsup X_n\right\}$$

$$=\bigcup_k\left\{\left(\liminf_{n\to\infty}X_n\ \ (\omega)\leq\frac{1}{k}\right)\bigcap\left(\limsup_{n\to\infty}X_n(\omega)\leq\frac{1}{k}\right)^c\right\}\in\Phi$$

故 $A\in\Phi$。

(2) 由命题 3.3，即得 $X(\omega)=\lim X_n\ \ (\omega)$ 是 $r.v$。

5. 随机变量的离散化逼近问题

本小节讨论对一般 $r.v$，如何用离散型 $r.v$(或称简单函数)序列的极限表示问题。

定义 3.3：设（Ω,Φ,P），$\{B_k,1\leq k\leq n\}$是Ω的一个有限分解($n\geq 1$), 及互不相同的实数 $\{a_k,1\leq k\leq n\}$, 称 $r.v.$ $X(\omega)=\sum_{k=1}^{n}a_kI_{A_k}(\omega)$ 为简单函数(或阶梯 $r.v$):

注：这里定义的简单函数 $X(\omega)$ 即为取值有限个的 $d.r.v$

命题 3.5 设 $P(X(\omega)<\infty)=1$,则 $X(\omega)$为（Ω,Φ）上 $r.v$ 的充要条件是存在简单函数序列 $\{X_n(\omega)\ \ ,n\geq 1\}$，满足：对每个 $\omega\in\Omega$ 有：$X(\omega)=\lim_{n\to\infty}X_n\ \ (\omega)$。

且当 $X(\omega)\geq 0$时,$\{X_n,n\geq 1\}$可取为非负递增简单函数序列。

证明：充分性，设 $\{X_n,n\geq 1\}$ 为简单函数序列,且 $\forall\omega\in\Omega$，$X(\omega)=\lim_{n\to\infty}X_n\ \ (\omega)$，由命题 3.4 知 $X(\omega)$是 $r.v$。

下证必要性：　先设 $X(\omega)\geq 0$，取

$$X_n(\omega)=\sum_{k=0}^{n.2^n-1}\frac{k}{2^n}I_{\{k\cdot 2^{-n}\le X(\omega)<(k+1)2^{-n}\}}+n\cdot I_{(X(\omega)\ge n)}$$

则 X_n 是简单函数，且在 $\{X<n\}$ 上，$0\le X-X_n\le\frac{1}{2^n}$，而在 $\{X\ge n\}$ 上，$X_n=n<X$，

当 $n\uparrow$ 时，显然 $X_n\uparrow$，故对每个 $\omega\in\Omega$。$X(\omega)=\lim_{n\to\infty}X_n(\omega)$。

对一般 $r.v.X(\omega)$，取 $X^+=X\vee 0, X^-=-(X\wedge 0)$，显然 $X^+,X^-\ge 0$，由上，对 X^+,X^-，

存在简单函数序列，$X_n^+\nearrow X^+$；$X_n^-\nearrow X^-$，取 $X_n=X_n^+-X_n^-$，则对每个 $\omega\in\Omega$，

$X_n(\omega)\to X(\omega)$

§4 连续型随机变量的条件概率密度

因为(X,Y)是连续型 $r.v.$，此时对任意 x,y，$P(X=x)=0$，$P(Y=y)=0$；因此不能直接用离散型 $r.v.$定义条件概率分布的方法来讨论 $c.r.v.$的条件分布。

首先给出条件是连续型 $r.v.$的条件分布函数的定义：

定义4.2 给定y，设$\forall\varepsilon>0$，$P(y-\varepsilon<Y\le y+\varepsilon)>0$，若$\forall x\in\mathrm{R}$

$$\lim_{\varepsilon\to 0^+}P\{X\le x|y-\varepsilon<Y\le y+\varepsilon\}=\lim_{\varepsilon\to 0^+}\frac{P\{X\le x,y-\varepsilon<Y\le y+\varepsilon\}}{P\{y-\varepsilon<Y\le y+\varepsilon\}}$$

存在，则称此极限为 X在 $Y=y$ 条件下的**条件分布函数**。简称 X关于 $Y=y$ 条件分布。记作：$F_{X|Y}(x|y)\hat{=}P(X\le x|Y=y)$。

定义 4.3 对于二维 $r.v.(X,Y)$，若存在非负函数 $f_{X|Y}(x|y)$，使 $\forall y\in R$，$\forall B\in$ ✍，有

$$P(X\in B|Y=y)=\int_{x\in B}f_{X|Y}(x|y)dx,$$

则称 $f_{X|Y}(x|y)$ 为 X 关于 $Y=y$ 的**条件概率密度函数**。简记为 $c.p.d.f.$

定理 4.1 设(X,Y)的 $j.p.d.f.$为 $f(x,y)$。若 $f(x,y),f_Y(y)$ 在点（x,y）及点 y 处分别连续，

且 $f_Y(y)>0$，则

$$f_{X|Y}(x|y)=\frac{f(x,y)}{f_Y(y)}。\tag{4.1}$$

证明：

$$P(X\le x\mid Y=y)=\lim_{\varepsilon\to 0^+}\frac{P\{X\le x,y-\varepsilon<Y\le y+\varepsilon\}}{P\{y-\varepsilon<Y\le y+\varepsilon\}}=\frac{\lim\limits_{\varepsilon\to 0^+}\int_{y-\varepsilon}^{y+\varepsilon}\int_{-\infty}^{x}f(u,v)dudv\Big/2\varepsilon}{\lim\limits_{\varepsilon\to 0^+}\int_{y-\varepsilon}^{y+\varepsilon}f_Y(v)dv\Big/2\varepsilon}$$

$$\overset{*}{=}\int_{-\infty}^{x}f(u,y)du/f_Y(y)=\int_{-\infty}^{x}\frac{f(u,y)}{f_Y(y)}du$$（上面带（*）等号是根据积分中值定理）

故 $f_{X|Y}(x|y)=\dfrac{f(x,y)}{f_Y(y)}$。

推论 1

$$f(x,y)=f_{X|Y}(x\mid y)f_Y(y)\tag{4.2}$$

推论 2　$\forall y\in R$

$$P(Y\le y)=\int_{-\infty}^{+\infty}P(Y\le y\mid X=x)f_X(x)dx\tag{4.3}$$

证明：

$$P(Y\le y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{y}f(u,v)dudv=\int_{-\infty}^{+\infty}\int_{-\infty}^{y}\mathrm{f}_{Y|X=x}(v|u)\mathrm{f}_X(u)dudv$$

$$=\int_{-\infty}^{+\infty}\left[\int_{-\infty}^{y}\mathrm{f}_{Y|X=u}(v|u)dv\right]\mathrm{f}_X(u)du=\int_{-\infty}^{+\infty}P(Y\le y\mid X=u)f_X(u)du$$

推论 2 是全概率公式的推广。

例 4.1 若 (X,Y)在$G=\{(x,y):x^2+y^2\le 1\}$上服从均匀分布，即$f(\mathrm{x},\mathrm{y})=\frac{1}{\pi}I_{\left(x^2+y^2\le 1\right)}$。

求：$f_{Y|X=x}(y|x)$。

解： 当 $|\mathrm{x}|\le 1$时 $f_X(x)=\int_{-\sqrt{1-\mathrm{x}^2}}^{\sqrt{1-\mathrm{x}^2}}\frac{1}{\pi}dy=\frac{2}{\pi}\sqrt{1-\mathrm{x}^2}$；当 $|\mathrm{x}|>1$时 $f_X(x)=0$

由定理6.1易得：$f_{Y|X=x}(y|x)=\dfrac{1}{2\sqrt{1-\mathrm{x}^2}}I_{(-\sqrt{1-x^2},\sqrt{1-x^2})}(y),(\forall|x|<1)$

例 4.2 设 *r.v.* $X\sim U(0,1)$。当观察到 $X=x$，（$0<x<1$）时，数 Y 是区间（x，1）上的均匀分布，求（X,Y）的 *j.p.d.f.*。

解：X 具有概率密度 $f_X(x)=I_{(0<x<1)}$

类似的，在 $X=x$ 条件下 Y 的概率密度 $f_{Y|X}(y|x)=\dfrac{1}{1-x}I_{(x<y<1)}$

由推论1： $f(x,y)=f_{Y|X}(y|x)f_X(x)=\dfrac{1}{1-x}I_{(0<x<y<1)}$

命题 4.1 设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则$\forall x\in R$：

$$f_{Y|X=x}(y|x)\sim N(\mu_2+\rho\tfrac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2))$$

证明：由(2.2)式(X,Y)的 *j.p.d.f.*为：

$$f(x,y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\cdot\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x-\mu_1)(x-\mu_2)}{\sigma_1\sigma_2}+\frac{(y-\mu_2)^2}{\sigma_2^2}\right]\right\}$$

及§2 命题 2.1 知

$$f_X(x)=\int_{-\infty}^{+\infty}f(x,y)dy=\frac{1}{\sqrt{2\pi}\sigma_1}e^{-\frac{(x-\mu_1)^2}{2\sigma_1^2}}$$

由（4.1）式，得

$$f_{Y|X=x}(y|x)=\frac{1}{\sqrt{2\pi}\sigma_2\sqrt{1-\rho^2}}\cdot\exp\left\{\frac{-1}{2\sigma_2^2(1-\rho^2)}\left[y-\left(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1)\right)\right]^2\right\}\qquad（4.4）$$

与一维正态概率密度函数比较，可知 $f_{Y|X=x}(y|x)\sim N(\mu_2+\rho\frac{\sigma_2}{\sigma_1}(x-\mu_1),\sigma_2^2(1-\rho^2))$

这说明二维正态的条件分布仍是正态分布。

§5 随机变量的独立性

在第一章中，我们讨论了事件的独立性。本章第一节给出离散型 *r.v.*的独立性。在此基础上,我们来讨论一般随机变量的独立性。

定义 5.1 设 *r.v.*(X,Y)，若对$\forall(x,y)\in R^2$，有 $P\{X\leqslant x,Y\leqslant y\}=P\{X\leqslant x\}P\{Y\leqslant y\}$，则称 X 和 Y 是**相互独立**的。

可见，*r.v.*(X,Y)相互独立，等价于：$\forall(x,y)\in R^2$，事件$\{X\leqslant x\}$与$\{Y\leqslant y\}$均相互独立。

由定义,不难证明以下命题.

命题 5.1　下列叙述等价：

（1）(X,Y)独立，

（2）$\forall(x,y)\in R^2$，$P(X\le x|Y\le y)=P(X\le x)$，

（3）$\forall(x,y)\in R^2$，$P(Y\le y|X\le x)=P(Y\le y)$。

证明：留给读者作为练习.

命题 5.2　若(X,Y)是具有概率密度函数的随机变量，则下列三个命题等价：

（1）X,Y 相互独立的；

（2）$\forall(x,y)\in R^2$，$f(x,y)=f_X(x)f_Y(y)$（几乎处处）；

（3）$\forall(x,y)\in R^2$，$f_{X|Y=y}(x\mid y)=f_X(x)$（几乎处处）。

证明：由 $p.d.f.$的定义有：$P\{X\le x,Y\le y\}=\int_{-\infty}^{x}\int_{-\infty}^{y}f(x,y)dxdy$　（5.1）

$$P(X\le x)P(Y\le y)=\int_{-\infty}^{x}f_X(x)dx\int_{-\infty}^{y}f_Y(y)dy=\int_{-\infty}^{x}\int_{-\infty}^{y}f_X(x)f_Y(y)dxdy \quad (5.2)$$

先证（1）$\Rightarrow$（2）

若 X, Y独立，即$\forall(x,y)\in R^2$，$P\{X\leqslant x,Y\leqslant y\}=P\{X\leqslant x\}P\{Y\leqslant y\}$；即（5.1）=（5.2）。

可得　$f(x,y)=f_X(x)f_Y(y)$　（几乎处处）

下证（2）$\Rightarrow$（1）$\forall(x,y)\in R^2$，若 $f(x,y)=f_X(x)f_Y(y)$（几乎处处）可得（5.1）=（5.2），即 $P\{X\leqslant x,Y\leqslant y\}=P\{X\leqslant x\}P\{Y\leqslant y\}$，所以 X,Y 相互独立。以上证明了（1）$\Leftrightarrow$（2）。

其它证明留给读者。

命题 5.3　设$(\mathrm{X},\mathrm{Y})\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，则：$X,Y$ 相互独立的充要条件是：$\rho=0$。

证明：由前面命题 2.1 可知

$$f_X(x)=\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1{}^2}\right\}\quad f_Y(y)=\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left\{-\frac{(y-\mu_2)^2}{2\sigma_2{}^2}\right\}$$

所以　$f_X(x)\cdot f_Y(y)=\dfrac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\dfrac{(x-\mu_1)^2}{2\sigma_1{}^2}-\dfrac{(y-\mu_2)^2}{2\sigma_2{}^2}\right\}$

而X,Y　相互独立的充要条件 是：$f(x,y)=f_X(x)f_Y(y)$，$\forall(x,y)\in R^2$。

即$\forall(x,y)\in R^2$

$$\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\bullet\exp\left\{\frac{-1}{2(1-\rho^2)}\left[\frac{(x-\mu_1)^2}{2\sigma_1^{\ 2}}+\frac{(y-\mu_2)^2}{2\sigma_2^{\ 2}}-2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2}\right]\right\}$$

$$=\frac{1}{2\pi\sigma_1\sigma_2}\exp\left\{-\frac{(x-\mu_1)^2}{2\sigma_1^{\ 2}}-\frac{(y-\mu_2)^2}{2\sigma_2^{\ 2}}\right\}$$

上式取 $x=\mu_1, y=\mu_2$ 得 $\sqrt{1-\rho^2}=1\Rightarrow\rho=0$；反之，若 $\rho=0$，易知 $\forall(x,y)\in R^2$，

$f(x,y)=f_X(x)f_Y(y)$。　　证毕。

例 5.2　一负责人到达办公室的时间 X 均匀分布在 8~12 时，即 $X\sim U[8,12]$，其秘书到达办公室的时间 Y 均匀分布在 7~9 时，即 $Y\sim U[7,9]$，且二人到达办公室的时间独立，求他们到达办公室的时间相差不超过 5 分钟（1/12 小时）的概率。

解：由题设 X,Y 的 $p.d.f$ 分别为：

$$f_X(x)=\frac{1}{4}I_{(8<x<12)}\qquad f_Y(y)=\frac{1}{2}I_{(7<y<9)}$$

记图 5.1 中阴影部分为 $G=(|x-y|\le 1/12)\bigcap(8\le x\le 12, 7\le y\le 9)$，其面积为 S_G。显然题意所求概率为 $P\{(X,Y)\in G\}$，而 S_G=1/6，故

$$P\{(X,Y)\in G\}=\iint_{(x,y)\in G}f(x,y)dxdy=\frac{S_G}{8}=\frac{1}{48}$$

即负责人和他的秘书到达办公室的时间相差不超过 5 分钟的概率为 1/48。

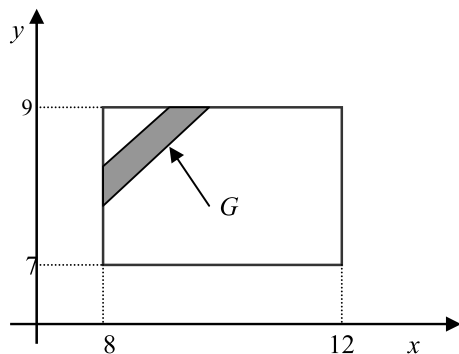


图 5-1

以上所述的关于二维随机变量的一些概率特性，容易推广到 n 维随机变量的情况。

定义 5.2　设 $X_1,\cdots,X_n$ 是概率空间 (Ω,Φ,P) 上的 n 个随机变量，若对 $\forall(x_1,\cdots,x_n)\in R^n$，有 $P(X_1\le x_1,\cdots,X_n\le x_n)=P(X_1\le x_1)\cdots P(X_n\le x_n)$，则称 $X_1,\cdots,X_n$ 相互独立。

定义 5.3　设 n 个随机变量 $X_1,\cdots,X_n$ 相互独立且服从同一分布，即

$$F_{X_i}(x)\triangleq F(x)，\forall i\in[1,n]$$

则称 $X_1,\cdots,X_n$ **独立同分布**，(*Independent Identically distributed*)简记作 ***i.i.d***。

对于 n 维随机变量，命题 5.2 可推广为：设连续型 *r.v.s.* $X_1,\cdots,X_n$ 的 *p.d.f* 分别为 $f_1(x),\cdots,f_n(x)$，则 $X_1,\cdots,X_n$ 相互独立的充要条件为：$f(x_1,x_2,\cdots,x_n)=\prod_{i=1}^{n}f_i(x_i)$。证明与命题 5.2 证明类似，从略。

命题 5.4　$X_1,\ldots,X_n$ 相互独立，则

（1）$\forall 1\le k<n$ $\sum_1^k X_i$与$\sum_{k+1}^n X_l$ 相互独立；$\sum_1^k X_i^2$与$\sum_{k+1}^n X_l^2$ 相互独立；

（2）若 $Y_i=g_i(X_i),i=1,\cdots,n$；即 Y_i 仅是 X_i 的函数，则 $Y_1,\cdots,Y_n$ 独立；

（3）若 $Z_1=h_1(X_1,X_2,\cdots,X_k)$，$Z_2=h_2(X_{k+1},\cdots,X_n),1\le k<n$,则 Z_1、Z_2 独立。

其中 $g_i(x),i=1,\cdots,n$、h_1 和 h_2 均为 *Borel* 可测函数。

证明：（1）的证明留给读者作为练习。（2）、（3）的证明已超出本书范围，从略。

例 5.3　设 X,Y，*i.i.d.*且$X\sim Ex(\lambda)$，而 $U=X^2$，$V=2Y$，求(U,V)的 *j.p.d.f*。

解： $P(U\le u)=P(X^2\le u)=P(0<X\le\sqrt{u})I_{(u>0)}=\int_0^{\sqrt{u}}e^{-x}dxI_{(u>0)}=\left(1-e^{-\sqrt{u}}\right)I_{(u>0)}$

$$f_U(u)=\frac{d}{du}P(U\le u)=e^{-\sqrt{u}}\cdot\frac{1}{2\sqrt{u}}I_{(u>0)}$$

$$P(V\le v)=P(2Y\le v)=P(0<Y<\frac{v}{2})I_{(v>0)}=\int_0^{\frac{v}{2}}e^{-y}dyI_{(v>0)}=(1-e^{-\frac{v}{2}})I_{(v>0)}$$

$$\therefore f_V(v)=\frac{d}{dv}P(V\le v)=\frac{1}{2}e^{-\frac{v}{2}}I_{(v>0)}$$

及定理 5.2：由 X,Y独立，得 U,V独立

$$\therefore f(u,v)=f_U(u)\cdot f_V(v)=\frac{1}{4\sqrt{\pi}}e^{-\left(\sqrt{u}+\frac{v}{2}\right)}I_{(u>0,v>0)}$$

定义 5.4 称定义在(Ω,Φ,P)上的随机变量序列$\{X_n,n\ge 1\}$相互独立，如果其中任意有限多个随机变量相互独立。

§6 随机向量函数的分布

1 随机向量函数的分布

这里主要讨论 *r.v.*的和、差、积、商等的分布。

（1）*d.r.v.*和的分布

由本章命题 1.1 知,两个 *d.r.v* 的和、差、积、商.极大极小仍为 *d.r.v*,那么如何求其分布律?举例如下.

例 6.1 (泊松分布对和的封闭性) 设 X_1，X_2 独立，$X_k \sim Po(\lambda_k)$,k=1,2；证明：$X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$。

证：因为 $\forall k \in N_0$，$(X_1 + X_2 = k) = \bigcup\limits_{i=0}^{k}(X_1 = i, X_2 = k-i)$；及 X_1，X_2 独立，故

$$P(X_1 + X_2 = k) = \sum_{i=0}^{k} P(X_1 = i)P(X_2 = k-i) = \sum_{i=0}^{k} \frac{\lambda_1^i \lambda_2^{k-i}}{i!(k-i)!} e^{-(\lambda_1+\lambda_2)}$$

$$= \frac{1}{k!}(\sum_{i=0}^{k} C_k^i \lambda_1^i \lambda_2^{k-i}) e^{-(\lambda_1+\lambda_2)} = \frac{(\lambda_1+\lambda_2)^2}{k!} e^{-(\lambda_1+\lambda_2)}$$

$\Rightarrow X_1 + X_2 \sim Po(\lambda_1 + \lambda_2)$。

（2） 不同类型的二个 *r.v.*的和（差、积、商）的分布

命题 6.1 设(X,Y)是(Ω, Φ)上的 *r.v*, X 为 *d.r.v*，则 $X \pm Y, X \cdot Y, X/Y(Y \neq 0)$ 均是 *r.v.*。

证明: 设 X 的分布律为：$P(X = x_i) = p_i$，则 $\forall a \in R$,

$$(X + Y \le a) = \bigcup_{i}(X = x_i, Y \le a - x_i) \in \Phi。$$

故 X+Y 也是 *r.v.*。

其它证明留给读者自己完成。

例 6.2 设 X,Y 独立，$X \sim N(\mu, \sigma^2)$，$P(Y = a_1) = p > 0$，$P(Y = a_2) = 1 - p = q > 0$；试问 X+Y 是否具有 *p.d.f.*？若有求 X+Y 的 *p.d.f.*。

解：注意到这是二个不同类型的 *r.v.*之和，Y 为 *d.r.v.*，X 为 *c.r.v.*，它们之和 X+Y 是什么类型的 *r.v.*呢？不妨看 $\forall z \in R, P(X + Y = z) = ?$ 显然

$$P(X+Y=z)=P(Y=a_1,X=z-a_1)+ P(Y=a_2,X=z-a_2)=0，$$

此时断定它可能是 *c.r.v.*，因而可先求其分布函数。注意到$\Omega=(Y=a_1)\bigcup(Y=a_2)$，

$(X+Y\le z)=(Y=a_1,X\le z-a_1)\bigcup(Y=a_2,X\le z-a_2)$；又因 *X*,*Y* 独立，所以

`$P(X+Y\le z)=P\left(X\le z-a_1\right)P\left(Y=a_1\right)+P\left(X\le z-a_2\right)P\left(Y=a_2\right)$

上式两边对 *z* 求导存在$(\forall z\in R)$，所以 *X*+*Y* 是 *c.r.v.*，且 *X* 是正态，故

$$f_{X+Y}(z)=P(Y=a_1)f_X(z-a_1)+P(Y=a_2)f_X(z-a_2)$$

$$=\frac{1}{\sqrt{2\pi}\sigma}\left[p\cdot\exp\left(-\frac{(z-a_1-\mu)^2}{2\sigma^2}\right)+q\cdot\exp\left(-\frac{(z-a_2-\mu)^2}{2\sigma^2}\right)\right]$$

说明：

（*i*）上式$f_{X+Y}(z)$是两个正态 *p.d.f.* $f_X(z-a_1)$、$f_X(z-a_2)$的“凸组合”，其形状如图6-1。这种 *p.d.f.*在工程上有着广泛的应用。有兴趣的读者试举这种 *p.d.f.*的实例。

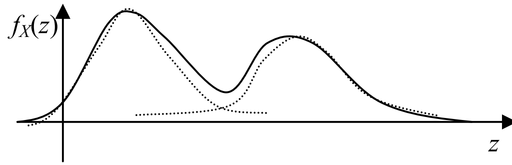


图 6-1

（*ii*）请读者思考：若 *X* 为 *c.r.v.*，*Y* 为 *d.r.v.*，如何求 *XY*，*X*/*Y*（*Y*≠0）的 *p.d.f.*。

命题 6.2 设(*X*,*Y*)是(Ω, Φ)上的 *r.v*，则$X\pm Y, X\cdot Y, X/Y(Y\neq 0)$均是 *r.v.*。

证明:由命题 3.5 知：存在 *d.r.v.*序列$\{X_n, n\ge1\}$使$X_n\to X$，由命题 6.1 知$\{X_n+Y, n\ge1\}$是 *r.v.*，显然$\forall\omega\in\Omega$有$X_n+Y\to X+Y$，再由命题 3.4 知 *X*+*Y* 也是 *r.v.*。

其它情形类似可证，留给读者作为练习。

（3）连续型 *r.v.*和（差、积、商）的分布

备忘录（预备知识）

为了方便读者阅读以下内容，这里列举分析中有关含参量积分和变上限积分的求导公式。

设 *g*(*x*,*t*)二元连续函数，且$\frac{\partial g}{\partial t}$连续，一元函数$\varphi(t)$关于 *t* 可导；记

$S_1(t)=\int_a^b g(x,t)dx$，$S_2(t)=\int_a^{\varphi(t)} g(x,t)dx$，则

$$\frac{dS_1(t)}{dt}=\int_a^b \frac{\partial g(x,t)}{\partial t}dx\text{，} \tag{6.1}$$

$$\frac{dS_2(t)}{dt}=\int_a^{\varphi(t)} \frac{\partial g(x,t)}{\partial t}dx+g(\varphi(t),t)\varphi'(t)\text{。} \tag{6.2}$$

例 6.3 (和的卷积公式)设(X,Y)的 *j.p.d.f.*为 $f(x,y)$，它关于(x,y)连续，求 $Z=X+Y$ 的分布及 *p.d.f.*。

解： $\forall z\in R$，利用全概率公式，易知 Z 的分布函数为

$$P(X+Y\le z)=\int_{-\infty}^{+\infty}P(Y\le z-x\mid X=x)f_X(x)dx \tag{6.3}$$

若 Y 关于 $X=x$ 的条件 *p.d.f.*为 $f_{Y|X}(y\mid x)$，由 $f(x,y)$连续，易知 $f_{Y|X}(y\mid x)$ 关于 y 连续，故由（6.3）式两边对 z 求导，得

$$f_{X+Y}(z)=\frac{dP(X+Y\le z)}{dz}=\int_{-\infty}^{+\infty}f_{Y|X=x}(z-x\mid x)f_X(x)dx \tag{6.3a}$$

故

$$f_{X+Y}(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx \tag{6.4}$$

同理可得

$$f_{X+Y}(z)=\int_{-\infty}^{+\infty}f(z-y,y)dy \tag{6.5}$$

若 X,Y 为非负 *c.r.v.*，则 $\forall z\ge 0$

$$f_{X+Y}(z)=\int_0^z f(x,z-x)dx \tag{6.6}$$

若 X,Y 独立，则

$$f_{X+Y}(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)dx \tag{6.7}$$

注：（6.7）式是 X,Y 独立时，*p.d.f.*的卷积公式，记作 f_X*f_Y。

若 X,Y 独立且非负，则 $\forall z\ge 0$

$$f_{X+Y}(z)=\int_0^z f_X(x)f_Y(z-x)dx \tag{6.8}$$

例 6.4 已知 X,Y 独立同分布且 $X\sim U(0,1)$，求 $X+Y$ 的 *p.d.f.*

解： 由题设 X,Y 独立。

由式(6.7)有$f_X(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)dx$

易知：当且仅当 $\begin{cases}0<x<1\\0<z-x<1\end{cases}$，亦即$\begin{cases}0<x<1\\z-1<x<z\end{cases}$时，上式不为零。

参照图6.2易得：

$$f_Z(z)=\begin{cases}0 & z\le 0或z>2\\ \int_0^z dx & 0<z\le 1\\ \int_{z-1}^1 dx & 1<z\le 2\end{cases}$$

得：$f_Z(z)=\begin{cases}0 & z\le 0或z>2\\ z & 0<z\le 1\\ 2-z & 1<z\le 2\end{cases}$

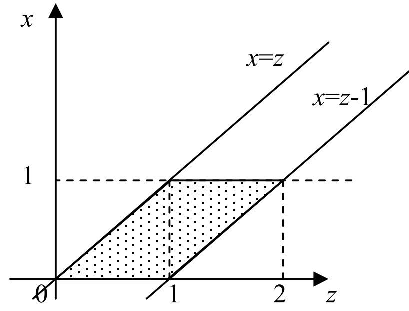


图 6-2

例 6.5 X,Y 互相独立，$X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，证明：$Z=X+Y$ 的 *p.d.f.*为正态概率密度，即：

$$X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$$

解：由式(6.7)可得：

$$f_Z(z)=\int_{-\infty}^{+\infty}f_X(x)f_Y(z-x)dx$$

$$=\int_{-\infty}^{+\infty}\frac{1}{\sqrt{2\pi}\sigma_1}\exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)\bullet\frac{1}{\sqrt{2\pi}\sigma_2}\exp\left(-\frac{(z-x-\mu_2)^2}{2\sigma_2^2}\right)dx$$

$$=\int_{-\infty}^{+\infty}\frac{1}{2\pi\sigma_1\sigma_2}\exp\left(-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)\bullet\exp\left(-\frac{\sigma_1^2+\sigma_2^2}{\sigma_1^2\sigma_2^2}\left(x-\frac{\sigma_2^2\mu_1+\sigma_1^2z-\sigma_1^2\mu_2}{\sigma_1^2+\sigma_2^2}\right)^2\right)dx$$

$$=\frac{1}{\sqrt{2\pi(\sigma_1^2+\sigma_2^2)}}\exp\left(-\frac{(z-\mu_1-\mu_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)$$

即 $X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2)$。

若 $(X,Y)\sim N(\mu_1,\sigma_1^2,\mu_2,\sigma_2^2,\rho)$ 其中 $\rho\ne 0$，即 X,Y 不独立，求 $Z=X+Y$ 的 *p.d.f.*，则有：

$$X+Y\sim N(\mu_1+\mu_2,\sigma_1^2+\sigma_2^2+2\rho\sigma_1\sigma_2)$$

为证上述结论，只需利用公式 $f_Z(z)=\int_{-\infty}^{+\infty}f(x,z-x)dx$ 及本章(2.2)式即可得到以上结

论。

若 $X_i \sim N\left(\mu_i, \sigma_i^2\right) \quad i=1,2,\ldots,n$，且它们相互独立，则它们的和 $Z=X_1+X_2+\ldots+X_n$ 仍然服从正态分布，且有：

$$\sum_{k=1}^{n} X_k \sim N\left(\sum_{k=1}^{n}\mu_k, \sum_{k=1}^{n}\sigma_k^2\right)$$

更一般的，可以证明，任意有限个正态随机变量（可以不独立）的线性组合仍然服从正态分布。

2 随机向量变换的分布

这里讨论如下问题：若 n 维 $r.v.$ $X=(X_k, 1\le k\le n)$ 的联合 $p.d.f.$ 为 $f_X(x)$，$x=(x_k, 1\le k\le n)\in R^n$，如何求 X 经变换（记为Λ）后的 n 维 $r.v.$ $Y=(Y_k=g_k(X), 1\le k\le n)$ 的分布或 $j.p.d.f.$。

设变换Λ：$y=(y_k=g_k(x), 1\le k\le n)\,\hat{=}\,(g(x))\in R^n, x\in R^n, g(x)=(g_k(x), 1\le k\le n)$

对应的逆变换Λ'：$x=(x_k=h_k(y), 1\le k\le n)\,\hat{=}\,(h(y))\in R^n, y\in R^n, h(y)=(h_k(y), 1\le k\le n)$

其中 g_k, h_k 是 n 元(*Borel* 可测（*注）)函数。

记 X 和 Y 的取值范围为 $G=\{x: f_X(x)>0, x\in R^n\}$，$G'=\{y: y=g(x), x\in G\}$。

$J=\det\left(\dfrac{\partial x_i}{\partial y_j}\right)_{n\times n}$ 为变换的雅可比(*Jacobi*)行列式。

命题 6.1 若上述变换Λ满足：

（1）变换Λ：$G\xrightarrow{g}G'$ 存在唯一的逆变换 ☹′：$G'\xrightarrow{h}G$；

（2）g_k, h_k 有连续偏导数（分别在 G 和 G' 上）；

（3）在 G 中，$J\ne 0$（几乎处处）；

则 n 维 $r.v.$向量 $Y=(Y_k=g_k(X), 1\le k\le n)\,\hat{=}\,(g(X))$ 的联合 $p.d.f.$ 为

$$f_Y(y)=f_Y(h(y))\,|J|, \quad y\in G' \tag{6.9}$$

证： $\forall D' \subset G', D' \in \mathscr{B}^2$，记 $D = \{x : x = h(y), y \in D'\}$，则 $D \subset G, D \in \mathscr{B}^2$（**注）；显然 $(Y \in D') = (X \in D)$，故

$$P(Y \in D') = P(X \in D) = \int_{x \in D} \cdots \int f_X(x)dx = \int_{y \in D'} \cdots \int f_X(h(y)) \,|\, J \,|\, dy$$

由 *j.p.d.f.*的定义知可取 $f_Y(y) = f_X(h(y)) \,|\, J \,|, y \in G'$ 作为 Y 的 *j.p.d.f.*。

（*注）所谓 $g(x)$为 n 维 *Borel* 可测函数，即 $\forall a \in R, \{x : g(x) \le a\} \in \mathscr{B}^n$（$B^n$ 为 n 维 *Borel* σ-域（代数））。

（**注）由 g,h 可测，则由 $D' \in \mathscr{B}^n$ 可得 $D \in \mathscr{B}^n$，于是 $(Y \in D') = (X \in D) \in \mathscr{F}$。

例 6.6　设(X,Y) *i.i.d.*，$X \sim N(0,1^2)$。令 $X = \rho\cos\theta, Y = \rho\sin\theta$，其中 $\rho \ge 0$，$0 \le \theta \le 2\pi$，试求 (ρ,θ) 的 *j.p.d.f.*及边缘 *p.d.f.*。

解： 令 $x = r\cos\alpha, y = r\sin\alpha$，得 $r = \sqrt{x^2 + y^2}$，$\alpha = tg^{-1}\frac{y}{x}$。又$(X,Y)$的 *j.p.d.f.*为

$$f(x,y) = \frac{1}{2\pi}\exp[-(x^2 + y^2)/2]$$

而 $J = \begin{vmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \alpha} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \alpha} \end{vmatrix} = r$，故 (ρ,θ) 的 *j.p.d.f.*为

$$f_{(\rho,\theta)}(r,\alpha) = \frac{1}{2\pi} \cdot re^{-r^2/2} \cdot I_{(r \ge 0, 0 \le \alpha \le 2\pi)}$$

而 ρ 的 *p.d.f.*为

$$f_\rho(r) = r \cdot e^{-r^2/2} \cdot I_{(r \ge 0)}$$

称此分布为瑞利(*Rayleigh*)分布。
θ的 *p.d.f.*为

$$f_\theta(\alpha) = \frac{1}{2\pi} \cdot I_{(0 \le \theta \le 2\pi)}$$

显而易见，$\forall(r,\alpha)$ 有 $f_\rho(r)f_\theta(\alpha) = f_{(\rho,\theta)}(r,\alpha)$，故 ρ,θ 独立。

说明：
（1）若(X,Y)表示射击弹着点的两个坐标，则 (ρ,θ) 表弹着点与原点的距离，θ为弹着点和原点连线与 x 轴的夹角。

（2）若(X,Y)*i.i.d.*，$X \sim N(0,\sigma^2)$ 则 ρ 是瑞利分布。

例 6.7 设 X,Y 相互独立，都服从参数 $\lambda=1$ 的指数分布，令 $U=X+Y$，$V=X/Y$；求(U,V)的 $j.p.d.f.$。

解：首先 (X,Y)的 $j.p.d.f$ 为：$f_{(X,Y)}(x,y)=e^{-(x+y)}I_{(x>0,y>0)}$

对变换 $\begin{cases} u= & x+y \\ v= & x/y \end{cases}$，显然满足命题的条件(1)、(2)、(3)；

解得：$\begin{aligned} x= & uv/(1+v) \\ y= & u/(1+v) \end{aligned}$ 求导可得：

$$J=\begin{vmatrix} \dfrac{v}{1+v} & \dfrac{u}{(1+v)^2} \\ \dfrac{1}{1+v} & -\dfrac{u}{(1+v)^2} \end{vmatrix}=-\frac{u}{(1+v)^2}$$

得： $f_{(U,V)}(u,v)=e^{-u}\cdot\dfrac{u}{(1+v)^2}I_{(u>0,v>0)}$。

§7 顺序统计量的分布

定义 7.1 给定 $(\Omega,☞,P)$，$\{X_k,1\le k\le n\}$ 为 n 维随机向量。$\forall\omega\in\Omega$，将 $X_1(\omega),\cdots,X_n(\omega)$ 按从小到大顺序排列，记为 $X_{(1)}(\omega)\le X_{(2)}(\omega)\le\cdots\le X_{(n)}(\omega)$，称 $X_{(1)}(\omega),\cdots,X_{(n)}(\omega)$ 为 $(X_1,\cdots,X_n)$ 的**顺序统计量**。

例：设一随机实验是每次投掷红白二颗骰子，记 $X_1(\omega)$, $X_2(\omega)$ 为相应的点数。如第一次试验结果是 $X_1(\omega_1)=5$, $X_2(\omega_1)=2$，则 $X_{(1)}(\omega_1)=2$, $X_{(2)}(\omega_1)=5$；如第二次结果是 $X_1(\omega_2)=1$, $X_2(\omega_2)=3$，则 $X_{(1)}(\omega_2)=1$, $X_{(2)}(\omega_2)=3$；$\propto$ 。

显然，$X_{(1)}=\min\limits_{1\le k\le n}X_k$；$X_{(n)}=\max\limits_{1\le k\le n}X_k$

称 $X_{(1)}$ 为极小量，$X_{(n)}$ 为极大量，$R_n=X_{(n)}-X_{(1)}$ 为极差。

顺序统计量在工程管理等领域有重要应用。本章仅限制在$\{X_k, 1\le k\le n\}$相互独立条件下讨论。

1　极大、极小分布

这里直接利用定义与事件转化的方法求之。

注意到$\forall z\in R$

$$(X_{(n)}\le z)=\bigcap_{k=1}^{n}(X_k\le z)，\ (X_{(1)}>z)=\bigcap_{k=1}^{n}(X_k>z)。$$

又由$\{X_k, 1\le k\le n\}$相互独立，得

$$P(X_{(n)}\le z)=\prod_{k=1}^{n}P(X_k\le z)，$$

$$P(X_{(1)}>z)=\prod_{k=1}^{n}P(X_k>z)，$$

进一步，若$\{X_k, 1\le k\le n\}$ *i.i.d.*，则

$$P(X_{(n)}\le z)=(P(X_1\le z))^n， \tag{7.1}$$

$$P(X_{(1)}>z)=(P(X_1>z))^n。 \tag{7.2}$$

以下是一个有趣结果

命题 7.1　若$\{X_k, 1\le k\le n\}$独立，$X_k\sim Ex(\lambda_k)\quad 1\le k\le n$，则$X(1)\sim Ex(\sum_{k=1}^{n}\lambda_k)$，即：

$$P(X_{(1)}\le z)=(1-e^{-(\sum_{k=1}^{n}\lambda_k)z})\cdot I_{(z\ge 0)}$$

证：由（7.2）式立得。

说明相互独立的指数分布其极小量仍是指数分布。

极大极小的背景：设n个电路元件的工作寿命为$X_1,\cdots, X_n$，（a）若将它们组成一个串联系统，且系统能正常工作是当且仅当n个元件均正常工作，于是该系统能正常工作的寿命为$X_{(1)}=\min\limits_{1\le k\le n}X_k$；（$b$）若将它们组成并联系统，且系统能正常工作是当且仅当至少有一元件正常工作，于是系统的工作寿命为$X_{(n)}=\max\limits_{1\le k\le n}X_k$。

有兴趣的读者试举出你生活中出现的极大极小的例子。

2 $(X_{(1)},X_{(n)})$ 的联合分布

这里仅在$\{X_k,1\le k\le n\}$ $i.i.d.$的情况下讨论。

为求$(X_{(1)},X_{(n)})$的联合分布$P(X_{(1)}\le x,X_{(n)}\le y),\forall(x,y)\in R^2$，我们直接用事件转化分解的方法求之。

分区域考虑，不妨设X_k与X同分布。

当$x\ge y$时，显然$(X_{(1)}\le x,X_{(n)}\le y)=(X_{(n)}\le y)$，故有

$$P(X_{(1)}\le x,X_{(n)}\le y)=P(X_{(n)}\le y)=(P(X\le y))^n\text{。}$$

当$x<y$时，$(x<X_{(1)},X_{(n)}\le y)=\bigcap_{k=1}^{n}(x<X_k\le y)$，有

$$P(x<X_{(1)},X_{(n)}\le y)=[P(X\le y)-(P(X\le x)]^n\text{。}$$

由 $(X_{(n)}\le y)=(X_{(1)}\le x,X_{(n)}\le y)\bigcup(x<X_{(1)},X_{(n)}\le y)$，

即 $(X_{(1)}\le x,X_{(n)}\le y)=(X_{(n)}\le y)-(x<X_{(1)},X_{(n)}\le y)$。

从而当$x<y$时，

$$P(X_{(1)}\le x,X_{(n)}\le y)=P(X_{(n)}\le y)-P(x<X_{(1)},X_{(n)}\le y)\text{。}$$

得

$$P(X_{(1)}\le x,X_{(n)}\le y)=[P(X\le y)]^n-[P(X\le y)-P(X\le x)]^n$$

所以

$$P(X_{(1)}\le x,X_{(n)}\le y)=\begin{cases}[P(X\le y)]^n & x\ge y,\\ [P(X\le y)]^n-[P(X\le y)-P(X\le x)]^n & x<y.\end{cases}$$

极差：$R_n=X_{(n)}-X_{(1)}$是刻划$X_1,\propto,X_n$取值波动大小的量。

思考：如何求R_n的分布？

下面讨论二种主要的顺序统计量的分布。

3 均匀分布顺序统计量的分布

设$\{X_k, 1 \le k \le n\}$ *i.i.d.*且$X_k \sim U[0,t], (t>0)$，$X_{(1)}, \propto, X_{(n)}$为其顺序统计量。

例 7.1　*n*=2 时，求$(X_{(1)}, X_{(2)})$的联合 *p.d.f.*。

解：设$f(x_1,x_2)$为$(X_{(1)}, X_{(2)})$的*j.p.d.f.*，用微元法求之。因$0 \le X_{(1)} \le X_{(n)} \le t$，显然当$f(x_1,x_2)$仅在 $0<x_1<x_2<t$ 区域非零，在其他区域上$f(x_1,x_2)=0$；故只需考虑 $0<x_1<x_2<t$ 的情形。取充分小 $h>0$ 满足 $0<x_1<x_1+h<x_2<x_2+h<t$　　(*)
则

$$\begin{aligned}&\{x_1 < X_{(1)} \le x_1+h, x_2 < X_{(2)} \le x_2+h\}\\&=\{x_1 < X_1 \le x_1+h, x_2 < X_2 \le x_2+h\} \cup \{x_1 < X_2 \le x_1+h, x_2 < X_1 \le x_2+h\}\end{aligned}$$

且由（*）式条件，知等式右边二事件不相容，故

$$P(x_1 < X_{(1)} \le x_1+h, x_2 < X_{(2)} \le x_2+h) = 2!\frac{h^2}{t^2}$$

因而在 $0<x_1<x_2<t$ 上

$$f(x_1,x_2) = \lim_{h\to 0}\frac{P(x_1 < X_{(1)} \le x_1+h, x_2 < X_{(2)} \le x_2+h)}{h^2} = \frac{2!}{t^2}$$

故$(X_{(1)}, X_{(2)})$的*j.p.d.f.*为

$$f(x_1,x_2) = \frac{2!}{t^2}\cdot I_{(0<x_1<x_2<t)}$$

类似于例 7.1 的证明，我们有

命题 7.2　设$\{X_k, 1 \le k \le n\}$ *i.i.d.*且$X_k \sim U[0,t], (t>0)$，则其顺序统计量的联合 *p.d.f.*为

$$f(x_1,\cdots,x_n) = \frac{n!}{t^n}\cdot I_{(0<x_1<\cdots<x_n<t)} \qquad (7.3)$$

证明留给读者。

作为练习，请读者在 *n*=3 时，给出上述命题的严格证明。并仔细考虑每步做法的必要性与根据。

例 7.2　*n*=2 时，（1）令$Y_1 = X_{(1)}, Y_2 = X_{(2)} - X_{(1)}$，求$(Y_1, Y_2)$的联合 *p.d.f.*；（2）问$Y_1, Y_2$是否独立？是否同分布？试证明你的猜想。

解：（从略，作为练习，留给有兴趣的读者完成。）

如读者做出例 7.2，则不难证明以下结果：（初次读本书的读者，可先跳过以下命题）

命题 7.3 设$\{X_k, 1\le k\le n\}$ $i.i.d.$且$X_k \sim U[0,t],(t>0)$，$X_{(1)},\propto,X_{(n)}$为其顺序统计量，则$X_{(1)},X_{(2)}-X_{(1)},\cdots,X_{(n)}-X_{(n-1)},t-X_{(n)}$不独立，但同分布。

证：从略，有兴趣的同学可作为练习。

4. 指数分布顺序统计量的分布

例 7.3 设X_1,X_2独立，且$X_k \sim Ex(\lambda_k),(1\le k\le 2)$，$X_{(1)}$，$X_{(2)}$为其顺序统计量，求（$X_{(1)},X_{(2)}$）的联合 $p.d.f.$。

解：设（$X_{(1)},X_{(2)}$）的联合 $p.d.f.$为$f(t_1,t_2)$；因$0\le X_{(1)}\le X_{(2)}$，只需考虑$0<t_1<t_2$的情形，用微元法类似于例 7.1 的讨论可得：

$$f(t_1,t_2)=\lambda_1\lambda_2(e^{-(\lambda_1 t_1+\lambda_2 t_2)}+e^{-(\lambda_1 t_2+\lambda_2 t_1)})I_{(0<t_1<t_2)}$$

当$\lambda_i=\lambda$时，$f(t_1,t_2)=2\lambda^2 e^{-\lambda(t_1+t_2)}I_{(0<t_1<t_2)}$

应用微元法，可证以下命题。

命题 7.4 设$\{X_k,1\le k\le n\}$ $i.i.d.$且$X_k\sim Ex(\lambda)$，$X_{(1)},\propto,X_{(n)}$为其顺序统计量，则

（1）（$X_{(1)},\propto,X_{(n)}$）的联合 $p.d.f.$为

$$f(t_1,\cdots,t_n)=n!\lambda^n e^{-\lambda(t_1+\cdots+t_n)}I_{(0<t_1<\cdots<t_n)} \qquad (7.4)$$

（2）$X_{(1)},X_{(2)}-X_{(1)},\cdots,X_{(n)}-X_{(n-1)}$独立，但不同分布。

（3）
$$X_{(1)}\sim Ex(n\lambda) \qquad (7.5)$$

$$X_{(k+1)}-X_{(k)}\sim Ex((n-k)\lambda)，\ 1\le k\le n-1 \qquad (7.6)$$

以上结果，有广泛的应用。参看第五章练习及第六章有关内容。

练习题

3.1 设 *d.r.v* (X,Y)的联合分布律同第 4 章例 5.1,求:

(1) 分别求 X 及 Y 的边缘分布律;

(2) 求 X 关于 $Y=i$ (i=1,…)的条件分布律;

(3) 求 $X+Y$ 的分布律;

(4) 求 XY 的分布律;

(5) 求 $X\vee Y$ 的分布律;

(6) 求 $X\wedge Y$ 的分布律;

(7) 求 $X\vee Y$ 与 $X\wedge Y$ 的联合分布律.

3.2 设（X，Y）的联合密度为：

$$f(x,y)=\frac{C}{(1+x^2)(1+y^2)}。$$

求（1）系数 C;

（2）（X，Y）落在以（0，0），（0，1），（1，0），（1，1）为顶点的正方形内的概率;

（3）X，Y 是否独立?

3.3 一机器制造直径为 X 的圆轴，另一机器制造内径为 Y 的轴衬，设（X，Y）的联合密度为：

$$f(x,y)=\begin{cases}2500, & 0.49<x<0.51，0.51<y<0.53 \\ 0, & 其他。\end{cases}$$

若轴衬内径与轴的直径之差在（0.004，0.036）内，则两者可以相适衬，求任一轴与任一轴衬相适衬的概率。

3.4 设(X,Y)的联合 *p.d.f.*是二元正态密度函数为：

$$f(x,y)=\frac{1}{2\pi}\exp\left(-\frac{1}{2}(2x^2+y^2+2xy-22x-14y+65)\right)。$$

（1）把它化为标准形式并指出 $\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho$；

（2）求出 $f_X(x)$ 及 $f_Y(y)$；

（3）求 Y 在 $X=x$ 时的条件 *p.d.f.*。

3.5 设系统 L 由两个相互独立的子系统 L_1，L_2 联接而成，联接的方式分别为（1）串联;（2）并联;（3）备用（当系统 L_1 损坏时，系统 L_2 即开始工作）。已知 L_1，L_2 的寿

命分别为 X 和 Y，相应的概率密度为：$f_X(x)=\alpha e^{-\alpha x}I_{(x>0)}$，

$f_Y(y)=\beta e^{-\beta y}I_{(y>0)}$，其中 $\alpha>0,\beta>0,\alpha\neq\beta$。试分别就这三种联接方式找出系统 L 的寿命 Z 的概率密度。

3.6 设某种商品一周的需要量是一个随机变量，其密度为：$f(x)=xe^{-x}I_{(x>0)}$。如果各周的需要量是相互独立的，试求：(1)两周的需要量的概率密度；(2)三周的需要量的概率密度。

3.7 若 X,Y 相互独立且有相同的分布 N（0，1），试证 $U=X^2+Y^2$ 与 $V=X/Y$ 相互独立。

3.8 已知 $X_1,\cdots,X_n$ 独立且同为参数为 λ 的指数分布，$X_{(1)},\ldots,X_{(n)}$为其顺序统计量：记

$$Y_1=X_{(1)},Y_2=X_{(2)}-X_{(1)}\cdots,Y_n=X_{(n)}-X_{(n-1)}$$

求证 $Y_1,\cdots Y_n$ 相互独立。

3.9 设 $X(\omega)\sim N$（$40,2^2$），令：

$$A=\{\omega:36\le X(\omega)\le 44\},Y(\omega)=\begin{cases}36, & \text{当}X(\omega)\le 36\\ X(\omega)\cdot I_A(\omega), & \\ 44, & \text{当}X(\omega)\ge 44\text{。}\end{cases}$$

$Z(\omega)=X(\omega)\wedge 44$。试求：（1）$Y(\omega)$ 及 $Z(\omega)$ 的分布函数。（2）$X(\omega)$ 在 A 发生的条件 $p.d.f.$。

3.10 设 $\{Y_n,n\ge 1\}i.i.d.$且$P(Y_n=1)=p>0,P(Y_n=-1)=1-p=q>0$，令 $X_0=0$，$X_n=\sum_{k=1}^{n}Y_k$。$A=\{X_3=1\},B=\{X_5-X_1=2\}$。（1）求 $P(AB),P(B\mid A)$；(2) 当 $p=2/3,q=1/3$ 时，A、B 是否相容？独立？证明你的结论。

3.11 设 $\{X_i,1\le i\le n\}$ 独立，X_i 的 $p.d.f.$ 为 $f_i(u),1\le i\le n,\xi$ 为 $d.r.v.$，分布为：$P(\xi=i)=p_i\ge 0,\sum_i p_i=1$，且 ξ 与 $\{X_i,1\le i\le n\}$ 独立。令 $X=\sum_{i=1}^{n}I_{(\xi=i)}\cdot X_i$，求 X 的 $p.d.f.$。试举出出现这个问题的实际例子。

3.12 设 $X_1,X_2,\cdots,X_k\ i.i.d.$，且 $X_i\sim G(p)$。求 $S_k=\sum_{i=1}^{k}X_i$ 的概率分布。

3.13 设 X,Y 独立，$X\sim N(\mu,\sigma^2),P(Y=1)=p>0,P(Y=2)=1-p=q>0$，求：

$Z=X\cdot Y$ 的 *p.d.f.* 。

3.14 设 X,Y 独立， $X\sim N(\mu,\sigma^2),Y$ 为离散型 *r.v.*， $P(Y=a_k)=p_k\geq 0,k\in N,$

$\sum\limits_k p_k=1$。求 $Z=X+Y$ 的 *p.d.f.*。

3.15 $\{Y_n,n\geq 0\}$ *i.i.d.*且P（$Y_n=1)=p>0,P(Y_n=0)=1-p=q>0,N\sim Po(\lambda)$ 且

与 $\{Y_n,n\geq 0\}$ 独立。试求 $S=\sum\limits_{n=0}^{N}Y_n$ 的分布律。

3.16 设 $X\sim N(10,2^2)$, $Y\sim B(10,1/2)$, X 与 Y 独立, 求 $Z=X+Y$ 的 *p.d.f.*。

3.17 设$(X,Y)\sim N(0,0,\sigma_1^2,\sigma_2^2,\rho)$，

(1)求 $Z=X/Y$的 *p. d. f.*；

(2)当$\rho=0$时, 证$Z=X/Y$的*p.d.f*为$f_2(\delta)=\alpha/\pi(\alpha^2+\delta^2)$,其中$\alpha=\sigma_1\sigma_2^{-1}$。

3.18 设 X,Y 独立， $X\sim N(\mu,\sigma^2),Y\sim G(p):P(Y=k)=(1-p)^{k-1}p,$

$k\in N$ 。试求： $W=X/Y$ 的 *p.d.f.*。

3.19 设 $\{X_i,1\leq i\leq n\}$ *i.i.d.*,$X_i\sim U(0,t),X_{(1)},X_{(2)},\cdots,X_{(n)}$ 为其顺序统计量，

$S_n=\sum\limits_{i=1}^{n}X_i$， (1)求 S_2 的 *p.d.f.*；(2)求 S_3 的 *p.d.f.*；(3)当 n=3 时，求

$X_{(1)},X_{(2)},X_{(3)}$ 的 *j.p.d.f.*；(4)当 n=3 时，试证： $X_{(1)},X_{(2)}-X_{(1)},X_{(3)}-X_{(2)},$

$t-X_{(3)}$ 不独立但同分布。

3.20 设 $\{X_i,1\leq i\leq n\}$ 独立，且 $X_i\sim Ex(\lambda_i)$　　$i=1,\cdots,n$ 。求：

(1) $X_1\wedge X_2,X_1\vee X_2,X_1+X_2$ 的 *p.d.f.*；

(2) $Z_n=X_1\wedge X_2\wedge X_3\wedge\cdots\wedge X_n$ 的 *p.d.f.*；

(3) X_1 在 $X_1\leq X_2$ 下的 *c.p.d.f.*；

(4) X_1 在 $X_1 \le X_2 \le X_3$ 下的 *c.p.d.f.*;

(5) X_2 在 $X_1 \le X_2 \le X_3$ 下的 *c.p.d.f.*。

3.21 设 $\{X_i, 1 \le i \le n\}$ 独立，$X_i \sim Ex(\lambda_i) \quad i = 1, \cdots, n$，$S_n = \sum_{i=1}^{n} X_i$，

$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 为其顺序统计量。(1)求 X_1 / X_2 的 *d. f.*；(2) 求 (S_1, S_2) 的 *j.p.d.f.*;(3)若 $\lambda_i = \lambda, 1 \le i \le n$, 试问: $X_{(1)}$，$X_{(2)} - X_{(1)}, \cdots, X_{(n)} - X_{(n-1)}$ 是否独立? 同分布? 试证你的猜想。

3.22 $\{X_i, 1 \le i \le n\}$ 独立同分布，$X_i \sim Ex(\lambda) \quad i = 1, \cdots, n$，$S_0 = 0$ $S_n = \sum_{i=1}^{n} X_i$ $\forall t \ge 0$，定义 $N(t) = \sum_{n=1}^{\infty} I_{(S_n \le t)}$。(1) 证 $N(1) \sim Po(\lambda)$；(2) $N(1)$ 与 $N(2) - N(1)$ 独立否? (3)求 S_1 在 $N(1) = 0$ 下的 *c.p.d.f.*;(4) S_1 在 $N(1) = 1$ 下的 *c.p.d.f.*。

(提示: $(N(1) = n) = (S_n \le 1 < S_{n+1})$)

3.23 设 $\{X_i, 1 \le i \le n\}$ 独立，$X_i \sim Ex(\lambda_i) \quad i = 1, \cdots, n$，$X_{(1)}, X_{(2)}, \cdots, X_{(n)}$ 为其顺序统计量，$\forall t > 0$，$N_n(t) = \sum_{i=1}^{n} I_{(X_{(i)} \le t)}, (n \ge 2)$。(1) 求 $N_3(t)$ 的分布律; (2) 当 n=2 时，求 $X_{(1)}$ 的 *p.d.f.* 及 $X_{(1)}$ 在 $X_1 \le X_2$ 下的条件 *p.d.f.*;(3)求 $X_{(1)}$ 在 $N_2(t) = 0$ 下的条件 *p.d.f.*；(4) 求 $X_{(2)}$ 在 $N_2(t) = 1$ 下的条件 *p.d.f.*。

3.24 记号和条件同 3.23。(1) 求 $X_{(1)}$，$X_{(2)}$ 在 $N_2(t) = 2$ 下的联合条件 *p.d.f.*;(2) 当 $\lambda_i = \lambda > 0, (1 \le i \le 2)$ 时，试问: $X_{(1)}$ 与 $X_{(2)} - X_{(1)}$ 是否条件独立? 同分布? 试证你的猜想。

3.25 设 $X_1, \cdots, X_n, i.i.d., X_k \sim N(\mu, 1^2)$, 令 $\overline{X} = \frac{1}{n} \sum_{k=1}^{n} X_k$,

$Y = \sum_{k=1}^{n} (X_k - \overline{X})^2$，(1) 求 $\overline{X}$ 的 *p.d.f.*；(2) 求 Y 的 *p.d.f.*，(3) 证明 $\overline{X}$ 和 Y 相互独立。

3.26 $\{Y_n, n \ge 1\} i.i.d. P(Y_n = 1) = p > 0, P(Y_n = 0) = r \ge 0, P(Y_n = -1) = q > 0$，

$p+r+q=1$。令 $X=0, X_n=\sum_{k=1}^{n} Y_k$, $A=(X_3=1), B=(X_5-X_1=2)$

（1）求 $P(A), P(B), P(AB), P(A|B),\ \ P(A\cup B)$；

（2）求 X_4 的分布律；

（3）求 X_4 在 $X_2\ge 0$ 下的条件分布律。

3.27 设 Y_n, X_n 同 3.26 中定义, 记

$T_{01}=\min\{n, n\ge 1, X_n=1\}, T=\min\{n, n\ge 1, X_n=-1$ 或 $X_n=2\}$

求(1) $P(T_{01}=k), P(T=k), (1\le k\le 6)$；(2)求 $(T\wedge 7)$ 的分布律及 T 在 $T\ge 7$ 下的条件分布律；（3）求 $P(X_T=-1)$ 及 $P(X_T=2)$。

3.28 设某种电缆有内外两层绝缘。记外层绝缘的寿命为 *X*, 内层绝缘的寿命为 *Y*，(*X,Y*) 有联合密度 $f(x,y)=2\theta^{-2}e^{-(x+y)/\theta}, 0<x<y<+\infty$。（1）设 *U=ln(X)*,求 *U* 的密度函数；(2)求已知外层绝缘寿命为 x 时, 内层绝缘的寿命的条件分布；(3) 求已知外层绝缘寿命大于 x_0 时, 内层绝缘的寿命小于 y_0 的条件概率。

3.29 设二维随机变量（*X,Y*）有如下分布：$\begin{pmatrix} (0,0) & (0,1) & (1,0) & (1,1) \\ p_1 & 0 & p_2 & p_3 \end{pmatrix}$

(1)求 *max*{*X,Y*}和 *min*{*X,Y*}的联合分布律；(2)给出 *X* 和 *Y* 独立的条件。

3.30 设 $X_1,\cdots,X_n, i.i.d., X_k \sim N(0,1^2), Y\sim N(0,1^2)$, 且与$\{X_k, 1\le k\le n\}$独立, 记

$\chi^2(n)=\sum_{k=1}^{n} X_k^{\ 2}, T(n)=Y/(\chi^2(n)/n)^{1/2}$。称 $\chi^2(n)$ 为自由度为 n 的 χ^2 分布；$T(n)$ 是自由度为 n 的 t−分布。它们在数理统计中有广泛应用。

求(1) $\chi^2(n)$ 的 *p.d.f.*；

(2) $T\ (n)$ 的 *p.d.f.*。

3.31 设 X_1, X_2 独立，且 $X_1=\lambda^2(n), X_2=\lambda^2(m)$，其中 $\lambda^2(n)$ 如上题定义，记

$F(n,m)=n^{-1}X_1/m^{-1}X_2$，　　求 $F(n,m)$ 的 *p.d.f* 。

注：$\chi^2(n), T(n), F(n,m)$ 在数理统计中是三个最常用的$r.v.$，它们的 $p.d.f$ 分别称为自由度 是n的卡方分布，自由度为n的$t-$分布及自由度为(n,m)的$F-$分布。

3. 32 设f_n, f是定义在（Ω, Φ）上的 r.v. (n>0)，证明:

$$\left\{\omega: \lim_{n\to\infty} f_n(\omega) = f(\omega)\right\} = \bigcap_{k=1}^{\infty} \bigcup_{N=1}^{\infty} \bigcap_{n=N}^{\infty} \left\{\omega: \left| f_n(\omega) - f(\omega) \right| < \frac{1}{k}\right\}$$

$$\left\{\omega: \lim_{n\to\infty} f_n(\omega) \neq f(\omega)\right\} = \bigcup_{k=1}^{\infty} \bigcap_{n=1}^{\infty} \bigcup_{n=N}^{\infty} \left\{\omega: \left| f_n(\omega) - f(\omega) \right| \geq \frac{1}{k}\right\}$$

3. 33 设(X,Y)是(Ω,Φ)上的$r.v.$,证明: $X+Y$, $X-Y$, XY, 及 X/Y （$Y\neq 0$) 均是$r.v.$ 。

3. 34 设$\{X_k,\ 1\leq k\leq n\}$ $i.i.d.$,且X_1与X同分布，其分布函数为$F(x)=P(X\leq x)$,（这在数理统计中称$r.v.X$为总体,$\{X_k,\ 1\leq k\leq n\}$为其样本），$X_{(1)}, X_{(2)}, \cdots X_{(n)}$为其顺序统计量。$\forall x\in R$，

记$\alpha_n(x) = \sum_{k=1}^{n} I_{(X_{(k)}\leq x)}$， $\widetilde{F}_n(x) = \alpha_n(x)/n$，通常称$\widetilde{F}_n(x)$为$X$的经验分布函数。

（1） 试求$\alpha_n(x)$的分布律；

（2） 试证：对$\forall x\in R$固定，$\sum_{k=0}^{n} (\frac{k}{n}) P(\widetilde{F}_n(x) = \frac{k}{n}) = F(x)$。

3. 35 设$(X,Y) \sim N(\mu_1, \sigma_1^2, \mu_2, \sigma_2^2, \rho)$，令$Z_1 = X - \mu_1$，$Z_2 = Y - \mu_2 - \rho\sigma_2\sigma_1^{-1}(X - \mu_1)$。

（1）求Z_2的$p.d.f.$，及（Z_1, Z_2）的联合$p.d.f$。

（2）证明Z_1与Z_2相互独立。

第四章 数字特征

前面我们介绍了可以完满表达随机变量概率性质的分布函数 $F(x)$。然而在理论与应用中，人们往往更关心随机变量的某些主要特征和相互关系，这是引出数字特征的原因之一。

数字特征是由随机变量的分布函数 $F(x)$确定的，是对 $F(x)$进行某种运算的结果，以便突出反映其主要特征及相互间的某种关系。本章着重介绍若干常用的数字特征。其中，数学期望和条件数学期望是最重要而又基本的两个。引入这些重要概念将使随机数学的研究内容与应用更加多彩多姿。

特别要提一下的是，由于条件数学期望能更精细和深入的刻划多个随机变量之间的关系，以及它的良好性质。因此，它已在现代随机数学理论的迅猛发展和广泛应用的舞台上扮演越来越重要的角色。

§1 数学期望

1 数学期望的定义

在随机变量的数字特征中，我们常常关心随机变量取值“平均”的大小，能够反映这个特征的就是随机变量的数学期望。

我们先来看一个例子：

例 1.1 某人独立射击 100 次，结果如下：

击中环数	8	9	10
相应次数	10	20	70

如何衡量该射手的射击水平？显然首先要看其每次击中的平均环数。

解： 所求的平均环数$=8\times\frac{10}{100}+9\times\frac{20}{100}+10\times\frac{70}{100}=9.6$

可见，100 次射击的平均环数等于击中的环数乘以该环数出现的频率求和。此数更集中反映射手的水平。

我们将上例抽象化，记 X 为击中的环数，是 $d.r.vX$ 其分布律为：$P(X=x_i)=p_i$。则 X 的平均值为：$\overline{X}=\sum_{i=1}^{\infty}x_i\cdot p_i$ 。

定义 1.1 若 $d.r.v\,X$ 的分布律为：$P(X=x_i)=p_i$，$i\in\subseteq$，且 $\sum_i |x_i|\cdot p_i<\infty$ ，则称

$$EX \triangleq \sum_i x_i \cdot p_i \qquad (1.1)$$

为 X 的**数学期望** (*Expectation*)，简称期望。

特别的，当随机变量的取值为：$P(X=x_i)=\frac{1}{n}$，$1\le i\le n$。则：$EX=\frac{1}{n}\sum_{i=1}^{n}x_i$，即是 x_i（$1\le i\le n$）的算术平均值。所以，数学期望是算术平均的推广，即加权平均值。故亦称为均值（*mean*）。

注意：定义中要求该级数绝对收敛，是为了保证该和数不随求和次序的改变而改变。当 $\sum_i |x_i|\cdot p_i$ 发散时，称 X 的数学期望不存在。

例1.2 若 $X\sim B(1,p)$，则 $EX=0\times(1-p)+1\times p=p=P(X=1)$，即 $EX=P(X=1)$。

在这里我们注意到数学期望和概率分布的关系，为此考察事件 $A\in$ ☞ 的示性函数为 $I_A(\omega)$ 的期望 $E(I_A(\omega))=P(A)$，等于事件 A 的概率。这就将数学期望和事件的概率联系起来了。因此，求某事件发生的概率就可以化为求它的示性函数的数学期望。可见引进数学期望将丰富概率论的研究内容。

例 1.3 若 $X\sim B(n,p)$，则：

$$EX=\sum_{k=0}^{n}k\cdot P(X=k)=\sum_{k=0}^{n}\frac{k\cdot n!}{k!(n-k)!}p^kq^{n-k}=\sum_{k=1}^{n}\frac{n(n-1)!\,p}{(k-1)!(n-1-(k-1))!}p^{k-1}q^{n-1-(k-1)}$$

$$=np\sum_{k'=0}^{n-1}\frac{(n-1)!}{k'!(n-1-k')!}p^{k'}q^{n-1-k'}=np(p+q)^{n-1}=np.$$

例 1.4 若 $X\sim Po(\lambda)$， 则：$EX=\sum_{k=0}^{\infty}k\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\lambda\cdot(\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!})\cdot e^{-\lambda}=\lambda$

在求 *r.v.*的期望中，常要用到如下结果：

（1）$\sum_{k=0}^{\infty}x^k=\frac{1}{1-x}$，$|x|<1$；（2）$\sum_{k=1}^{\infty}kx^{k-1}=\frac{1}{(1-x)^2}$，$|x|<1$，（3）$\sum_{k=0}^{\infty}\frac{x^k}{k!}=e^x$。

因为当$|x|<1$ 时，$\sum_{k=1}^{\infty}kx^{k-1}=(\sum_{k=0}^{\infty}x^k)'_x=(\frac{1}{1-x})'=\frac{1}{(1-x)^2}$。

例 1.5 若 $X\sim G(p)$， 则：$EX=\sum_{k=1}^{\infty}k\cdot(1-p)^{k-1}\cdot p=\frac{1}{p}$

例 1.6 若 $X\sim Po(\lambda)$， 求：EX^2。

解：由于 X 为非负整数，此时 $P(X^2=k^2)=P(X=k)$，故：

$$EX^2=\sum_{k=0}^{\infty}k^2\cdot\frac{\lambda^k}{k!}\cdot e^{-\lambda}=\sum_{k=1}^{\infty}k\cdot\frac{\lambda^k}{(k-1)!}\cdot e^{-\lambda}=\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!}\cdot e^{-\lambda}+\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}\cdot e^{-\lambda}=\lambda^2+\lambda$$

由例 1.6 自然会想到，对于 X 的一般函数 $Y=g(X)$是否会有类似的结果？事实上，可以证明，对于离散型随机变量 X，$P(X=x_k)=p_k$，若 $g(x)$ 是连续函数或逐段单调函数，则 $Y=g(X)$也是离散型随机变量。且 $EY=E(g(X))=\sum_k g(x_k)\cdot P(X=x_k)$。

特殊情形：设 X 为取值非负整数 $r.v.$，则有：

$$EX=\sum_{n=1}^{\infty}P(X\geq n)$$

证明：留作练习。这里仅提示一下，注意应用下式：

$$EX=\sum_{n=1}^{\infty}nP(X=n)=\sum_{n=1}^{\infty}(\sum_{k=1}^{n}1)P(X=n)\text{。}$$

对于连续型随机变量（$c.r.v.$），有如下

定义 1.2 设 $c.r.v.\ X$ 的 $p.d.f$ 是 $f(x)$，若 $\int_{-\infty}^{\infty}|x|\cdot f(x)dx<\infty$，则称

$$EX\triangleq\int_{-\infty}^{\infty}x\cdot f(x)dx \qquad (1.2)$$

为 X 的**数学期望**。

例 1.7 若 $X\sim U[a，b]$，即 $f(x)=\frac{1}{b-a}\mathrm{I}_{(x\in[a,b])}$ 则： $EX=\int_a^b x\cdot\frac{1}{b-a}dx=\frac{a+b}{2}$

例 1.8 若 $X\sim Ex(\lambda)$，即 $f(x)=\lambda\cdot e^{-\lambda x}\mathrm{I}(x>0)$ 则： $EX=\int_0^{\infty}x\cdot\lambda e^{-\lambda x}dx=\frac{1}{\lambda}$

例 1.9 若 $X\sim N(\mu,\sigma^2)$ 则；

$$\begin{aligned}EX&=\frac{1}{\sqrt{2\pi}\cdot\sigma}\cdot\int_{-\infty}^{\infty}(x-\mu+\mu)\cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}dx\\&=\frac{1}{\sqrt{2\pi}\cdot\sigma}\left\{\int_{-\infty}^{\infty}(x-\mu)\cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}dx+\int_{-\infty}^{\infty}\mu\cdot e^{-\frac{(x-\mu)^2}{2\cdot\sigma^2}}dx\right\}=\mu\end{aligned}$$

等式右边第一项中被积函数为奇函数，故积分为 0。

前面分别对离散型和连续型两种随机变量定义了数学期望，对于一般随机变量，严格定义要用到黎曼-斯蒂尔吉斯积分，下面先介绍它的定义。

2 黎曼-斯蒂尔吉斯（*Reimann-Stieltjes*）积分

定义：设 $g(x),F(x)$ 是 R 上的函数（不妨限制在 $F(x)$为单调函数的情形），$\forall[a,b]$，设 $a=x_0<x_1<\cdots<x_n=b$，$\Delta x_k=x_k-x_{k-1}$，$\delta_n=\max\limits_{1\leq k\leq n}\Delta x_k$；任取 $u_k\in[x_{k-1},x_k]$，若极限 $\lim\limits_{\delta_n\to 0}\sum_{k=1}^{n}g(u_k)(F(x_k)-F(x_{k-1}))=I$ 存在，则称极限：

$$\int_a^b g(x)dF(x) \triangleq \lim_{\delta_n \to 0} \sum_{k=1}^{n} g(u_k)(F(x_k) - F(x_{k-1}))$$

为 $g(x)$ 关于 $F(x)$ 在$[a,b]$上的**黎曼-斯蒂尔吉斯积分**。(R-S)

若 $\lim\limits_{\substack{a\to-\infty \\ b\to+\infty}} \int_a^b g(x)dF(x) = \int_{-\infty}^{+\infty} g(x)dF(x)$ 存在且有限，则称此极限值为 $g(x)$ 关于 $F(x)$ 在 $(-\infty,+\infty)$ 上的**黎曼-斯蒂尔吉斯积分**。(简称 R-S 积分)

注意：

(1) 当取 $F(x)=x$ 时，则 $\int_{-\infty}^{+\infty} g(x)dF(x) = \int_{-\infty}^{+\infty} g(x)dx$ 化为通常的黎曼积分。

(2) 当取 $g(x)=x$，$F(x)$ 是 $d.r.v.X$ 的分布函数时，即若 $P(X=x_k)=p_k, k\in N$，则 $F(x)=\sum\limits_{x_k\le x} p_k$ 为台阶形跳跃函数，在 x_k 有跳跃高度 p_k，$k\in N$；则由定义可知：$\int_{-\infty}^{+\infty} xdF(x) = \sum\limits_k x_k p_k = EX$。

(3) 当取 $g(x)=x$，而 $F(x)$ 是 $c.r.v.X$ 的分布函数，X 的 $p.d.f.$ 为 $f(x)$ 或在 R 上几乎处处有 $F'(x)=f(x)$ 存在时，则可证明：$\int_{-\infty}^{+\infty} xdF(x) = \int_{-\infty}^{+\infty} xf(x)dx = EX$。此时上式左边为 x 关于 $F(x)$ 的 R-S 积分，而右边是函数 $xf(x)$ 的 *Riemann* 积分。

(4) R-S 积分与普通定积分有相类似的性质，如线性性，积分区间迭加性等。值得注意的是：若 a 点是 $F(x)$ 的跳跃点，则 $\int_{a^-}^{a^+} g(x)dF(x) = g(a)(F(a^+) - F(a^-))$。

(5) 设 $F(x)$ 是 $r.v.$ X 的分布函数，即 $F(x)=P(X\le x)$。则 $\forall a,b\in R, B\in \mathcal{B}$，$\int_a^b dF(x) = P(X\in[a,b]), P(X\in B) = \int\limits_{x\in B} dF(x)$

定义 1.3 若随机变量 X 的分布函数为 $F(x)$，且满足 $\int_{-\infty}^{\infty} |x| dF(x) < \infty$，则称

$$EX \triangleq \int_{-\infty}^{\infty} xdF(x) \qquad (1.3)$$

为 X 的**数学期望**。

对于随机变量函数的数学期望我们有：

定理 1.1 设 $g(x)$ 是一元波雷尔可测函数，随机变量 X 的分布函数为 $F(x)$，若 $\int_{-\infty}^{\infty} |g(x)| dF(x) < \infty$，则 $g(X)$ 的数学期望为：

$$Eg(X) = \int_{-\infty}^{\infty} g(x)dF(x) \qquad (1.4)$$

注：在第二章已提过，逐段单调函数、连续函数均为 *Borel* 可测函数。

严格证明这一定理需要用到测度论，已超出本教材的范围，这里不作要求，但要学会运用这一结论去推导其他性质。例如：

对于离散型随机变量 X，有：

$$E(g(X))=\sum_k g(x_k)\cdot P(X=x_k). \tag{1.5}$$

对于连续型随机变量 X，*p.d.f.*为 $f(x)$有：

$$Eg(X)=\int_{-\infty}^{\infty}g(x)\cdot f(x)dx 。\tag{1.6}$$

有了这样的性质，我们在求 $g(X)$的数学期望时，就可以不必求出 $g(X)$的分布，而只需要知道 X 的分布就可以了。若将 $g(X)$特殊化，就可以得到各种数字特征－－矩。有关矩的概念将在后面具体介绍。

特殊情形，设 X 为非负 *r.v.*，则：

$$EX=\int_0^{\infty}P(X>u)du$$

证明：留作练习。仅提示：

$$EX=\int_0^{\infty}xdP(X\le x)=\int_0^{\infty}(\int_0^{x}dy)dP(X\le x)=\int_0^{\infty}\int_y^{\infty}dP(X\le x)dy$$

3 数学期望的性质

数学期望有一些简单性质，我们列在下面，以下 X，Y 均表示随机变量。

（1）对于常数 C，$E(C)=C$，$E(X+C)=E(X)+C$；

（2）$E(CX)=CE(X)$，$E(C_1X+C_2Y)=C_1E(X)+C_2E(Y)$；

（3）若 X，Y 独立，则 $E(XY)=EX\,EY$；

（4）设 $g(x,y)$为波雷尔可测函数，X, Y 为连续型随机变量，其联合概率密度函数为 $f(x,y)$，则 $g(X,Y)$的数学期望为：

$$Eg(X,Y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}g(x,y)\cdot f(x,y)dxdy \tag{1.7}$$

我们仅对性质 3 在（X,Y）具有 *j.p.d.f.*时，给出证明如下：

设（X,Y）的 *j.p.d.f.*为 $f(x,y)$，$X\sim f_X(x)$，$Y\sim f_Y(y)$，则

$$(X,Y)\text{独立}\Leftrightarrow\forall(x,y)\in R^2，f(x,y)=f_X(x)f_Y(y)$$

所以

$$E(XY)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf(x,y)dxdy=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xyf_X(x)f_Y(y)dxdy$$

$$=\int_{-\infty}^{\infty}xf_X(x)dx\int_{-\infty}^{\infty}yf_Y(y)dy=EX\cdot EY$$

数学期望的定义在本章以后的章节中有基本的重要作用。掌握数学期望定义的确切含义及其直观意义，对以后概念的理解均有很大的帮助。

下面，我们就从结构角度理解数学期望的意义。

我们知道，数学期望：

$$EX=\begin{cases}\sum_k x_k \cdot P(X=x_k) & X\text{为}d.r.v.\\ \int_{-\infty}^{\infty} x\cdot f(x)dx & X\text{为}c.r.v\text{ 且具有}p.d.f: f(x)\end{cases}$$

以离散型随机变量 X 为例。前面我们定义了示性函数 $I_A(\omega)$，则：

$X(\omega)=\sum_k x_k \cdot I_{(X=x_k)}(\omega)$，因此：$EX=\sum_k x_k \cdot P(X=x_k)$。

样本空间Ω可以用 x_k 不同取值划分为若干个不相容事件的并集，EX 即为 X 在这些不相容事件上取值的加权平均；连续型随机变量也是一样，它可以表示成示性函数的线性组合的极限（见第二章），对 X 取值的加权平均就得到了以上的积分形式。

一个实际变量 X 可以写成 $X=EX+(X\text{-}EX)$，若 X 是测量值，则（$X\text{-}EX$）表示测量误差，其中 EX 是真值，相当于测量中我们真正想要的东西，即测量对象的精确值，而测量值总在真值 EX 附近波动，这是由于实际现象中的误差及其它偶然因素造成的。

§2 方 差

随机变量另一个重要的数字特征是方差，它反映了随机变量取值的集中与分散程度。在实际问题中，研究随机变量与其均值的偏离程度是十分重要的，通常我们用 $E[(X\text{-}EX)^2]$ 来度量这个偏离程度。

定义 2.1 设 X 是 $r.v.$，则称

$$DX=E[(X\text{-}EX)^2] \tag{2.1}$$

为 X 的方差（*Variance*），也可记作 $VarX$；称 $\sigma_X=\sqrt{DX}$ 为 X 的标准差或均方差。

注：

（1）DX 大，反映 X 取值比较分散；DX 小，反映 X 取值比较集中。

（2）常用的计算方差的公式为：

$$DX=EX^2-(EX)^2 \tag{2.2}$$

事实上：$DX=E(X-EX)^2=E(X^2-2XEX+(EX)^2)=EX^2-(EX)^2$.

在实际测量中，方差反映了 $r.v.$取值与数学期望的离散程度，用以刻画测量精度。DX 小，则说明测量精度高。

例 2.1 若 $X\sim B(1,p)$，则 $EX=p$，$EX^2=p$，$DX=EX^2-(EX)^2=p-p^2=pq$。

例 2.2 若 $X\sim B(n,p)$，则：$EX=np$，$EX^2=n(n\text{-}1)p^2+np$，$DX=npq$。

例 2.3 若 $X\sim P(\lambda)$， 则：$EX=\lambda$，$EX^2=\lambda^2+\lambda$，$DX=\lambda$。

例 2.4 若 $X\sim U[a,b]$，则：$EX=\frac{a+b}{2}$，$EX^2=\int_a^b x^2\cdot\frac{1}{b-a}dx=\frac{1}{3}(b^2+ab+a^2)$

$$DX=EX^2-(EX)^2=\frac{1}{12}(b-a)^2。$$

例 2.5 若 $X\sim E(\lambda)$，则：$EX=\frac{1}{\lambda}$　$DX=\frac{1}{\lambda^2}$。

例 2.6 若 $X\sim N(\mu,\sigma^2)$，则：$EX=\mu$；试求 DX

解：$DX=E(X-EX)^2=E(X-\mu)^2=\int_{-\infty}^{\infty}(x-\mu)^2\cdot\frac{1}{\sqrt{2\pi}\sigma}\cdot e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$。作变换 $t=\frac{x-\mu}{\sigma}$，则 $DX=\frac{\sigma^2}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2e^{-\frac{t^2}{2}}dt$，应用分部积分法，容易验算 $\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}t^2e^{-\frac{t^2}{2}}dt=1$。因此 $DX=\sigma^2$。

由例 2.6 我们知道，正态随机变量的概率密度函数中的两个参数 μ 和 σ^2 分别是该随机变量的数学期望和方差。因而，正态随机变量的分布完全可由它的数学期望和方差所决定。

注意到若 $X\sim N(\mu,\sigma^2)$，则 $P(\mu-3\sigma<x\le\mu+3\sigma)=\Phi(3)-\Phi(-3)=0.9974$，$P(|X-\mu|\le 2\sigma)=0.9545$。因此，对于正态随机变量，其值落在区间 $[\mu-3\sigma，\mu+3\sigma]$ 内的概率接近于 1。这也就看出了正态随机变量的集中程度。在工程测量中通常取 3σ 为误差限（精度要求高的），也有取 2σ 为误差限的（一般精度要求）。

方差的一些简单性质如下：

（1）$X=C$，或 $P(X=C)=1$，则：$DX=0$；

（2）$D(C_1X+C_2)=C_1^2D(X)$；

（3）对于随机变量 X,Y，有

$D(X\pm Y)=DX+DY\pm 2E[(X-EX)(Y-EY)]$

$D(aX\pm bY)=a^2DX+b^2DY\pm 2abE[(X-EX)(Y-EY)]$

若 X、Y 独立，则：$D(X\pm Y)=DX+DY$

证明：$D(X\pm Y)=E\{[(X\pm Y)-E(X\pm Y)]^2\}=E\{[(X-EX)\pm(Y-EY)^2]\}$

$=E[(X-EX)^2+(Y-EY)^2\pm 2(X-EX)(Y-EY)]$

$=E[(X-EX)^2]+E(Y-EY)^2\pm 2E[(X-EX)(Y-EY)]$

若 X、Y 独立，则 $X-EX$、$Y-EY$ 独立，由数学期望性质可知：

$E[(X-EX)(Y-EY)]=E(X-EX)\cdot E(Y-EY)=0$，故：$D(X\pm Y)=DX+DY$

（4）契比雪夫（*Chebyshev*）不等式：若 $DX<\infty$，则对任意 $\varepsilon>0$,

$$P(|X-EX|\ge\varepsilon)\le\frac{DX}{\varepsilon^2}\tag{2.3}$$

证明：设 $X\sim F(x)$

左 $= \int\limits_{|X-EX|\geq\varepsilon} dF(x) \leq \int\limits_{|X-EX|\geq\varepsilon} \frac{|X-EX|^2}{\varepsilon^2} dF(x) \leq \int_{-\infty}^{+\infty} \frac{|X-EX|^2}{\varepsilon^2} dF(x) = \frac{DX}{\varepsilon^2}$

契比雪夫不等式作为一个理论工具，其应用是很普遍的。这个不等式给出了在未知随机变量 X 的分布下，事件{ω：|X-μ|< ε}的概率的一种估计方法。

（5）DX=0，充要条件是：存在常数μ，使 $P(X=\mu)=1$。

证明："⇒"由 DX=0，及契比雪夫不等式有：$\forall n>0, P(|X-EX|\geq\frac{1}{n})\leq 0 \Rightarrow$

$\forall n>0, P(|X-EX|<\frac{1}{n})=1$，又 $(X=EX)=(\bigcap\limits_{n=1}^{+\infty}|(X-EX)|<\frac{1}{n})=\lim\limits_{n\to\infty}(|(X-EX)|<\frac{1}{n})$

故：$P(X=EX)=\lim\limits_{n\to\infty}P(|(X-EX)|<\frac{1}{n})=1.$

"⇐"若存在常数μ，使 $P(X=\mu)=1$。$\Rightarrow F_X(x)=I_{(X\geq\mu)}(x)\Rightarrow EX=\int_{-\infty}^{\infty}xdF_X(x)=$

$\int_{\mu^-}^{\mu}xdF_X(x)=\mu\Rightarrow DX=E(X-\mu)^2=\int_{\mu^-}^{\mu}(x-\mu)^2dF_X(x)=0.$

由性质 3 的证明可得：

（6）若 $E[(X-EX)(Y-EY)]\neq0$，则 X,Y 不独立。

这说明 $E[(X-EX)(Y-EY)]\neq0$，反映了 X,Y 之间有某种关系。

§3 协方差和相关系数

对于二维随机向量，要了解各个分量之间的关系，只有数学期望和方差是不够的，我们还需要引进反映它们之间关系的数字特征。

由上节方差的性质 6 可知：$E[(X-EX)(Y-EY)]\neq0$，可得 X,Y 不独立，因此可用它来刻划二者之间的某种关系程度。

1. 定义

定义 3.1：设 X, Y 为两个 *r.v.*，则称：

$$Cov(X, Y)=E[(X-EX)(Y-EY)] \quad (3.1)$$

为 X, Y 的**协方差**。称：

$$\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{DX}\cdot\sqrt{DY}} \quad (3.2)$$

为 X, Y 的**相关系数**。

相关系数是刻划 X, Y 的线性关系是否密切的尺度。

对于一般的随机变量，可以对其作一定的变换使其标准化。事实上，由期望和方差

的性质我们知道：$E(\frac{X-EX}{\sqrt{DX}})=\frac{1}{\sqrt{DX}}E(X-EX)=\frac{1}{\sqrt{DX}}(EX-EX)=0,$

$D(\frac{X-EX}{\sqrt{DX}})=\frac{1}{DX}D(X-EX)=\frac{1}{DX}DX=1$。称随机变量

$$\frac{X-EX}{\sqrt{DX}} \tag{3.3}$$

为**随机变量 *X* 的标准化**。而

$$Cov(\frac{X-EX}{\sqrt{DX}},\frac{Y-EY}{\sqrt{DY}})=E(\frac{X-EX}{\sqrt{DX}}\cdot\frac{Y-EY}{\sqrt{DY}})=\frac{Cov(X,Y)}{\sqrt{DX}\cdot\sqrt{DY}}=\rho_{XY}$$

因此，X, Y 的相关系数，即 X, Y 标准化后的协方差是一个没有量纲的量。

由定义，易知：$Cov(X,Y)=E(XY)-(EX)(EY)$，故：

$$\rho_{XY}=\frac{E(XY)-(EX)(EY)}{\sqrt{DX\cdot DY}} \tag{3.4}$$

定义 3.2 若 X,Y 之间相关系数 $\rho_{XY}=0$，则称 X,Y **线性不相关**。

2. 性质

协方差和相关系数具有如下一些性质：

（**1**）$Cov(X, Y)= Cov(Y, X)$，$Cov(X+c, Y+d)= Cov(X, Y)$，其中 c,d 为任意常数；

（**2**）$Cov(aX, bY)=abCov(X, Y)$，其中 a,b 为任意常数；

（**3**）$Cov(X_1+X_2, Y)= Cov(X_1, Y)+ Cov(X_2, Y)$，$Cov(\sum_{i=1}^{n}X_i,\sum_{j=1}^{m}Y_j)=\sum_{i=1}^{n}\sum_{j=1}^{m}Cov(X_i,Y_j)$；

以上性质(1)~(3)的证明留给读者。

（**4**）若 X，Y 独立，则：$Cov(X, Y)=0$，$\rho_{XY}=0$，即 X,Y 不相关；

证明：若 X，Y 独立，则由数学期望性质：$E(XY)=(EX)(EY)$及(3.4)式有：$Cov(X,Y)=E(XY)-(EX)(EY)= (EX)(EY) -(EX)(EY)=0$，从而：$\rho_{XY}=0$。

（**5**）*Schwarz* 不等式：若 $EX^2<\infty$，$EY^2<\infty$，则：

$$(EXY)^2\le EX^2\cdot EY^2 \tag{3.5}$$

$$Cov(X,Y)\le\sqrt{DX}\cdot\sqrt{DY} \tag{3.6}$$

证明：因 $\forall t\in R$，$(X-tY)^2\ge0$，故 $u(t)=E(X-tY)^2\ge0$ 恒成立。即：$E(Y^2t^2)-2E(XYt)+E(X^2)\ge0$ 恒成立。故二次三项式 $u(t)$的判别式：$\varDelta=4\cdot[E(XY)]^2-4\cdot EX^2\cdot EY^2\le0$。(3.5)式得证。

由(3.5)式有：

$$(Cov(X,Y))^2=(E(X-EX)(Y-EY))^2\le E(X-EX)^2\cdot E(Y-EY)^2=DX\cdot DY$$

即：$Cov(X,Y)\le\sqrt{DX}\cdot\sqrt{DY}$。

（6）$|\rho_{XY}|\le 1$；

证明：由(3.6)式即得。

（**7**）对于随机变量 X, Y，下列命题等价：

1）$\rho_{XY}=0$；

2）$Cov(X,Y)=0$；

3）$E(XY)=EX\cdot EY$；

4）$D(X\pm Y)=DX+DY$。

这些均是与不相关等价的命题，详细的证明留给读者作为练习。

（**8**）$|\rho_{XY}|=1 \Leftrightarrow P(\frac{Y-EY}{\sqrt{DY}}\pm\frac{X-EX}{\sqrt{DX}}=0)=1$

证明：$|\rho_{Y}|=1 \Leftrightarrow 2(\rho_{XY}\pm 1)=0 \Leftrightarrow \quad D(\frac{Y-EY}{\sqrt{DY}}\pm\frac{X-EX}{\sqrt{DX}})=D(\frac{Y-EY}{\sqrt{DY}})+$

$D(\frac{X-EX}{\sqrt{DX}})\pm 2E\left\{(\frac{Y-EY}{\sqrt{DY}})(\frac{X-EX}{\sqrt{DX}})\right\}=1+1\pm 2\rho_{XY}=0$，由§2 方差性质（5）知以上等

价于：$P(\frac{Y-EY}{\sqrt{DY}}\pm\frac{X-EX}{\sqrt{DX}}=0)=1$。

可见，当$|\rho_{XY}|=1$ 时，随机变量 X, Y 以概率 1 取值在一条直线上，即：

$$P(\omega: Y-EY=\mp\sqrt{\frac{DY}{DX}}(X-EX))=1.$$

由性质(3)知，若 X, Y 独立，则 X, Y 不相关。然而，X,Y 不相关，二者却不一定独立。但是，在一种重要的特殊场合即二维正态分布下，不相关和独立则是等价的。

例 3.1 若$(X,Y)\sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则：

(1) $Cov(X,Y)=\rho\sigma_1\sigma_2$，$\rho_{XY}=\rho$；

(2) X,Y 不相关 $\Rightarrow X,Y$ 独立。

解：由前已知：$EX=\mu_1, EY=\mu_2, DX=\sigma_1^2, DY=\sigma_2^2$，而：

$$Cov(X,Y)=\frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_1)\cdot(y-\mu_2)\cdot\exp\{-\frac{(x-\mu_1)^2}{2\sigma_1^2}\}$$
$$\cdot\exp\{\frac{-1}{2(1-\rho^2)}\cdot(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1})^2\}dxdy$$

若令 $t=\frac{1}{\sqrt{1-\rho^2}}\cdot(\frac{y-\mu_2}{\sigma_2}-\rho\frac{x-\mu_1}{\sigma_1}),\quad u=\frac{x-\mu_1}{\sigma_1}$ 则：

$$Cov(X,Y)=\frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(\sigma_1\sigma_2\sqrt{1-\rho^2}tu+\rho\sigma_1\sigma_2u^2)\cdot\exp(-\frac{u^2}{2}-\frac{t^2}{2})dtdu$$

$$=\frac{\sigma_1\sigma_2\sqrt{1-\rho^2}}{2\pi}(\int_{-\infty}^{\infty}u\exp(-\frac{u^2}{2})du)(\int_{-\infty}^{\infty}t\exp(-\frac{t^2}{2})dt)$$

$$+\frac{\rho\sigma_1\sigma_2}{2\pi}(\int_{-\infty}^{\infty}u^2\exp(-\frac{u^2}{2})du)(\int_{-\infty}^{\infty}\exp(-\frac{t^2}{2})dt)$$

$$=\frac{\rho\sigma_1\sigma_2}{2\pi}\cdot\sqrt{2\pi}\sqrt{2\pi}=\rho\sigma_1\sigma_2$$

则 $\rho_{XY}=\frac{Cov(X,Y)}{\sqrt{DX}\cdot\sqrt{DY}}=\rho$. 可知参数ρ就是 X,Y 的相关系数。

由第三章命题(5.3)知：(X,Y)服从二维正态分布，则 X,Y 独立□ ρ=0 □ X,Y 不相关。

下面我们通过讨论最佳线性预测（估计）问题，进一步揭示相关系数所刻划的概率意义。

3. 最佳线性均方预测（估计）。

对于两个随机变量 X, Y。记：$\mu_1=EX,\mu_2=EY,\sigma_1^2=EX>0,\sigma_2^2=EY>0,$

$\rho_{XY}=\rho$，若 X 可观测，而 Y 不可观测，希望用 X 的线性函数 $bX+a$ 作为 Y 的预测（或估计），记作：$\hat{Y}=bX+a$。称 $\delta=Y-\hat{Y}=Y-(bX+a)$ 为预测误差，称 $Q(a,b)=E\delta^2=E(Y-(bX+a))^2$ 为均方预测误差。问题是：如何选取 a, b，使其均方预测误差 $Q(a, b)$ 达最小！（如存在的话）。

命题 3. 1 当

$$b^*=\rho\sigma_2\sigma_1^{-1} \tag{3.7}$$

$$a^*=\mu_2-\rho\sigma_2\sigma_1^{-1}\mu_1 \tag{3.8}$$

$$\hat{Y}^*=b^*X+a^* \tag{3.9}$$

时，满足：$Q(a^*,b^*)=\min_{a,b}Q(a,b)$。

证明： $Q(a,b)=E[Y-(bX+a)]^2=E\{[Y-\mu_2-b(X-\mu_1)]+(\mu_2-b\mu_1-a)\}^2$

注意到：$E[Y-\mu_2-b(X-\mu_1)]=0$，故有：

$$Q(a,b)=E[Y-\mu_2-b(X-\mu_1)]^2+(\mu_2-b\mu_1-a)^2$$
$$=\sigma_2^2-2b\rho\sigma_1\sigma_2+b^2\sigma_1^2+(\mu_2-b\mu_1-a)^2$$
$$=\sigma_2^2(1-\rho^2)+\sigma_1^2(b-\rho\sigma_2\sigma_1^{-1})^2+(\mu_2-b\mu_1-a)^2 \qquad *$$

注意到上式右边第 2，3 项均非负，故 $\forall a,b\in R^2$ 有：$Q(a,b)\geq\sigma_2^2(1-\rho^2)$，于是选取

$b^*=\rho\sigma_2\sigma_1^{-1}$，$a^*=\mu_2-b^*\mu_1=\mu_2-\rho\sigma_2\sigma_1^{-1}\mu_1$ 时，(*)式右边第 2，3 项为 0，即：

$$Q(a^*,b^*)=\sigma_2^2(1-\rho^2)=\min_{a,b}Q(a,b) \qquad (\Delta)$$

定义：称 $\hat{Y}^*\doteq b^*X+a^*=\rho\sigma_2\sigma_1^{-1}(X-\mu_1)+\mu_2$ 为 Y 的最佳线性均方预测（估计）。

由命题 3.1 证明的最后等式（Δ）可知：

命题 3.2： $Q(a^*,b^*)=\sigma_2^2(1-\rho^2)$

(3.10)

由(3.10)式可看出，对于给定的 Y，用 $|\rho_{XY}|=|\rho|$ 越接近于 1 的 X，其 b^*X+a^* 越接近于 Y（均方意义下）。这表明 $|\rho|$ 的大小刻划了 X, Y 之间线性关系密切的程度。

§4 矩、协方差矩阵及 *n* 维正态分布

前面我们介绍的数学期望、方差、协方差等是随机变量最常用的数字特征，它们的自然推广是矩。矩有以下几种：

（1）原点矩：对任意 $k\in Z^+$，称 EX^k 为 X 的 k 阶原点矩。数学期望是一阶原点矩。

（2）中心矩：对任意 $k\in Z^+$，称 $E[X-EX]^k$ 为 X 的 k 阶中心矩。方差是二阶中心矩。

（3）混合矩（对多维分布而言）：对任意 k、$l\in Z^+$，若 EX^kY^l 存在，称它为 X,Y 的 $k+l$ 阶混合原点矩；若 $E[(X-EX)^k(Y-EY)^l]$ 存在，称它为 X,Y 的 $k+l$ 阶混合中心矩。协方差 $Cov(X,Y)$ 是 X,Y 的二阶混合中心矩。

例 4.1 若 $X\sim N(0,\sigma^2)$，则 X 的 $2k$（$k\in N$）阶矩：$EX^{2k}=(2k)!\sigma^{2k}/2^k\cdot k!$。（可以通过分部积分求得，留给读者作为练习。）特别的，当 k=2 时，$EX^4=3\sigma^4$。

协方差矩阵定义：设 n 维随机变量 $X=(X_1, X_2, \cdots X_n)^T$，若 $\sigma_i^2\doteq DX_i, 1\leq i\leq n$ 存在，记 $\sigma_{ij}=Cov(X_i, X_j)$，则称矩阵：

$$\Sigma \hat{=} E[(X-EX)(X-EX)^T]=(\sigma_{ij})_{n\times n}$$

为 n 维随机变量 $X=(X_k, 1\le k\le n)^T$ 的协方差矩阵。

显然，协方差矩阵是一个对称矩阵。易证它是非负定矩阵。当$\sigma_i>0$ 时，它是正定矩阵，且对角线之元素 $\sigma_{ii}=\sigma_i^2=DX_i, 1\le i\le n$。

例 4.2 设 $(X,Y)\sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$，则有：

$$\Sigma=\begin{bmatrix}\sigma_1^2 & \rho\sigma_1\sigma_2\\ \rho\sigma_1\sigma_2 & \sigma_2^2\end{bmatrix} \tag{4.1}$$

记：$x=(x_1, x_2,)^T$，$\mu=(\mu_1, \mu_2)^T$，则 $X=(X_1, X_2)^T$ 的 *j.p.d.f.*可表为：

$$f(x)=\frac{1}{(2\pi)^{\frac{2}{2}}|\Sigma|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\}$$

其中$|\Sigma|=\det\Sigma$。

这样，记二维正态随机变量 $X=(X_1, X_2)^T\sim N(\mu,\Sigma)$。

例 4.3 设 $X=(X_1, X_2)^T\sim N(\mu,\Sigma)$，令 $Y=AX$，其中 $A=(a_{ij})_{2\times 2}$ 为非退化 2 阶矩阵，则：

$$Y\sim N(A\mu, A\Sigma A^T)$$

即二维正态分布进行了线性变换之后仍为正态分布。

证明留给读者作为练习。

以下介绍多元正态分布随机向量。记：$\mu=(\mu_k, 1\le k\le n)^T, x=(x_k, 1\le k\le n)^T$ 为 n 维随机向量，$\Sigma=(\sigma_{ij})$ 为 n 阶正定对称矩阵。定义 n 元函数

$$f(x)=\frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\exp\{-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\} \tag{4.2}$$

命题：(1) $\forall x\in R^n$， $f(x)>0$， (2) $\int_{R^n} f(x)dx=1$。

即由(4.2)式定义的 $f(x)$是 n 维 *p.d.f.*

证明：(1)显然。以下证(2)。由Σ 为正定对称方阵，则存在非奇异矩阵 L，使 $\Sigma=LL^T$。令线性变换 $y=L^{-1}(x-\mu)$，其逆变换为 $x=Ly+\mu$。此变换的雅可比行列式为：$|L|=|\Sigma|^{\frac{1}{2}}$。故：

$$\int_{R^n} f(x)dx = \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\int_{R^n}\exp\{-\frac{1}{2}y^T y\}|\Sigma|^{\frac{1}{2}}dy$$
$$= \frac{1}{(2\pi)^{\frac{n}{2}}|\Sigma|^{\frac{1}{2}}}\int\cdots\int_{R^n}\exp\{-\frac{1}{2}\sum_{k=1}^{n}y_k^2\}dy_1\cdots dy_n$$
$$= \left\{\frac{1}{\sqrt{2\pi}}\int_R e^{-u^2/2}du\right\}^n = 1.$$

定义：若 n 维 *r.v.* $X=(X_k, 1\le k\le n)^T$ 的 *j.p.d.f.*是用(4.1)式定义的 $f(x)$，则称 X 为 n 维正态 *r.v.*，$f(x)$为 n 维正态概率密度函数。

n 维正态分布有许多优良性质将在第五章进一步讨论。

§5 条件数学期望

条件数学期望是随机数学中最基本、最重要的概念之一。在这一节中我们将引入条件数学期望的概念，并说明它的性质和应用。为了直观地对此概念有个正确的理解，我们先从离散型随机变量入手,再讨论连续性随机变量, 最后推广到多元随机变量的情形。

1 以离散型随机变量作为条件的条件数学期望

（1）定义与例子

定义 5.1 设 *d.r.v* X, Y，分布律为：$P(X=x_i, Y=y_j)$，若对$\forall j\in N$，$P(Y=y_j)>0$，我们称：

$$E(X|Y=y_j)=\sum_i x_i P(X=x_i|Y=y_j) \qquad (5.1)$$

为 X 关于 $Y=y_j$ 时的条件数学期望。

这是一种狭义的条件数学期望。现比较(无条件)数学期望 $E(X)=\sum_i x_i P(X=x_i)$ 与条件数学期望 $E(X|Y=y_j)$ 的异同点。$E(X)$ 是对所有的 $\omega\in\Omega$，$X(\omega)$取值全体的加权平均，而 $E(X|Y=y_j)$是 X 取值局限在 $\omega\in\{\omega：Y=y_j\}=B_j$ 的加权平均。可以这样理解：记 $B_j=\{\omega:Y=y_j\}$, $A_i=\{\omega:X=x_i\}$，于是 $\{A_i, i\in N\}$ 及 $\{B_j, j\in N\}$ 分别是样本空间 Ω的分解，当 $A_iB_j=\phi$时，$P(X=x_i, Y=y_j)=0$，于是：

$$E(X|Y=y_j)=\sum_i x_i P(X=x_i|Y=y_j)=\sum_{i\in D_j^1} x_i P(X=x_i|Y=y_j)$$

其中 $D_j^1=\{i:A_iB_j\neq\phi\}$，这正说明了 $E(X|Y=y_j)$是$\omega\in B_j$时$X(\omega)$ 的局部加权平均。

显然 $E(X|Y=y_j)$依赖于$Y=y_j$,即依赖于$\omega\in B_j=\{\omega:Y=y_j\}$。从全局看，我们有必要引入一个新的随机变量$E(X|Y)$，使它在$\omega\in B_j$(即$Y=y_j$)时，取值为：$E(X|Y=y_j)$，称$r.v.E(X|Y)$ 为$r.v.X$关于$r.v.Y$ 的条件数学期望。它亦是 $r.v.$ Y 的函数。确切定义如下：

定义 5.2 设 $d.r.v.$ X, Y, 分布律为：$P(X=x_i,Y=y_j)$，若对$\forall j\in N$，$P(Y=y_j)>0$, 则称

$$E(X|Y)=\sum_{j\in N}E(X|Y=y_j)\cdot I_{(Y=y_j)}(\omega). \tag{5.2}$$

为 $r.v.$ X 关于 Y 的**条件数学期望**。

上述 $E(X|Y)$ 定义包含如下的直观意义：

（1）$r.v.E(X|Y)$ 是ω的函数。当$\omega\in\{\omega:Y=y_j\}$时，$E(X|Y)$的取值为 $E(X|Y=y_j)$。事实上，它是局部平均$\{E(X|Y=y_j),j\in N\}$的统一表达式。因而条件数学期望更深入的描述了 X,Y 之间的某种关系。

（2）当 $E(X|Y=y_j)\neq E(X|Y=y_k)(j\neq k)$ 时，$P\{E(X|Y)=E(X|Y=y_j)\}=P(Y=y_j)$否则，令 $D_j=\{k:E(X|Y=y_j)=E(X|Y=y_k)\}$，则：

$$P\{E(X|Y)=E(X|Y=y_j)\}=\sum_{k\in D_j}P(Y=y_k)$$

这个直观意义为我们提供了求条件数学期望分布律的简单方法。

（3）$E(X|Y)$亦是 Y 的函数，由§1 中随机变量函数的期望公式，知它的数学期望

$$E\{E(X|Y)\}=\sum_j E(X|Y=y_j)P(Y=y_j)$$

下面举一个例子来加深对上述概念的理解和认识。

例 5.1：$r.v.(X,Y)$的联合分布律如下表所示，试求 $E(X|Y)$的分布律，$E(X)$, $E\{E(X|Y)\}$。

Y \ X	1	2	3	P_j

1	$\frac{2}{27}$	$\frac{4}{27}$	$\frac{1}{27}$	$\frac{7}{27}$
2	$\frac{5}{27}$	$\frac{7}{27}$	$\frac{3}{27}$	$\frac{15}{27}$
3	$\frac{1}{27}$	$\frac{2}{27}$	$\frac{2}{27}$	$\frac{5}{27}$
P_i	$\frac{8}{27}$	$\frac{13}{27}$	$\frac{6}{27}$	1

解： 为求 $E(X\mid Y=j)$，先求 $E(X=i\mid Y=j)$,其中 i,j=1,2,3

$$P\{X=1\mid Y=1\}=\frac{P(X=1,Y=1)}{P(Y=1)}=\frac{\frac{2}{27}}{\frac{7}{27}}=\frac{2}{7}$$

同理： $P\{X=2|Y=1\}=\frac{4}{7},P\{X=3\mid Y=1\}=\frac{1}{7}$

$$E(X\mid Y=1)=\sum_{i=1}^{3}x_iP(X=x_i\mid Y=1)=1\times\frac{2}{7}+2\times\frac{4}{7}+3\times\frac{1}{7}=\frac{13}{7}$$

同理：$E(X\mid Y=2)=\sum_{i=1}^{3}iP(X=i\mid Y=2)=1\times\frac{5}{15}+2\times\frac{7}{15}+3\times\frac{3}{15}=\frac{28}{15}$

$$E(X\mid Y=3)=\sum_{i=1}^{3}iP(X=i\mid y=3)=1\times\frac{1}{5}+2\times\frac{2}{5}+3\times\frac{2}{5}=\frac{11}{5}$$

由前述直观意义（2），可知 $P\{E(X\mid Y)=E(X\mid Y=j)\}=P(Y=j), j=1,2,3$。故 $E(X|Y)$ 的分布律列表如下：

$E(X\|Y)$	$\frac{13}{7}$	$\frac{28}{15}$	$\frac{11}{5}$
$P\{E(X\|Y)=E(X\|Y=j)\}=P(Y=j)$	$\frac{7}{27}$	$\frac{15}{27}$	$\frac{5}{27}$

于是 $E\{E(X\mid Y)\}=\frac{13}{7}\times\frac{7}{27}+\frac{28}{15}\times\frac{15}{27}+\frac{11}{5}\times\frac{5}{27}=\frac{52}{27}$

而 $EX=\sum_{i=1}^{3}iP(X=i)=1\times\frac{8}{27}+2\times\frac{13}{27}+3\times\frac{6}{27}=\frac{52}{27}$

可见，$E\{E(X|Y)\}=EX$，事实上，若(X,Y)为 $d.r.v$，且 $E|X|<\infty$时， $E\{E(X|Y)\}=E(X)$ 普遍成立，即局部加权平均后的加权平均等于总体的加权平均。

2. 条件数学期望的性质：

下面列出 $d.r.v.$的条件数学期望的若干性质，以下均假定 $E|X|<\infty$，$E|X_i|<\infty, 1\le i\le n$。

（1） $E(X)=E\{E(X|Y)\}$；

（2） $E(\sum_{i=1}^{n}C_iX_i\,|\,Y)=\sum_{i=1}^{n}C_iE(X_i\,|\,Y)$；

（3）若 X,Y 独立，则 $E(X|\,Y)=EX$;

（4）$E(X\,|X)=X$，$E\{g(X)|\,X\}\}=g(X)$;

（5）$\forall g(X),h(Y)$,若 $E|g(X)h(Y)|<\infty$，则 $E\{g(X)h(Y)\ \ |Y\}\}=h(Y)E\{g(X)|Y\}$；其中 g,h 是 *Borel* 可测函数。

这些性质的证明请读者思考并作为练习。

对性质（1），可作如下表述：

$$E(X)=E\{E(X|Y)\}=\sum_j E(X\,|\,y=y_j)P(Y=y_j)$$

我们看到这样的形式与全概率公式具有相似之处，是全概率公式的推广。

一般地，设 $D\in$B，X 关于事件$\{\omega:Y(\omega)\in D\}=\{Y\in D\}$的条件数学期望为 $E(X\,|\,Y\in D)\triangleq\sum_i x_iP(X=x_i\,|\,Y\in D)$。若 Y 为 $d.r.v$，而 X 为一般 $r.v.$，设 $P(Y=y_j)>0$，对 $\forall x\in R$，称 $F_{X|Y=y_j}(x\,|\,y_j)=\dfrac{P(X\le x,Y=y_i)}{P(y=y_j)}$ 为 X 关于 $Y=y_j$ 条件分布函数。

定义 5.3 在上述记号下，称 $E\,(\,X\,|Y=y_j)=\int_{-\infty}^{+\infty}xdF_{X|Y=y_j}(x\,|\,y_j)$ 为 X 关于 $Y=y_j$ 的**条件数学期望**；称 $E(X|Y)=\sum_j E(X\,|\,Y=y_j)I_{(Y=y_j)}$ 为 X 关于 Y 的**条件数学期望**。

下面通过两个例题来加深对性质的理解。

例 5.2 设 N_1,N_2 独立，$N_i\sim P(\lambda_i),i=1,2$；求 $E(N_1\,|\,N_1+N_2)$ 及 $E(N_1+N_2\,|\,N_1)$ 的分布律。

解：（1）$E(N_1|\,N_1+N_2)$ 的分布律：$\forall n\in N$。

$$E(N_1|N_1+N_2=n)=\sum_{k=0}^{n}k\cdot P(N_1=k\,|\,N_1+N_2=n\,)$$

而 $P(N_1=k\,|\,N_1+N_2=n)=\dfrac{P(N_1=k,N_1+N_2=n)}{P(N_1+N_2=n)},(0\le k\le n)$，其中：

$$P(N_1=k,N_1+N_2=n)=P(N_1=k,N_2=n-k)=P(N_1=k)\cdot P(N_2=n-k)=\frac{\lambda_1^k}{k!}\cdot\frac{\lambda_2^{n-k}}{(n-k)!}\cdot e^{-(\lambda_1+\lambda_2)},$$

$$P(N_1+N_2=n)=\sum_{l=0}^{n}P(N_1+N_2=n\,|\,N_2=l)\cdot P(N_2=l)$$

$$=\sum_{l=0}^{n}P(N_1=n-l\,|\,N_2=l)\cdot P(N_2=l)$$

$$=\sum_{l=o}^{n}(\frac{\lambda_1^{n-l}}{(n-l)!})\cdot(\frac{\lambda_2^{l}}{l!})\cdot e^{-(\lambda_1+\lambda_2)}=\frac{(\lambda_1+\lambda_2)^n}{n!}\cdot e^{-(\lambda_1+\lambda_2)}$$

故

$$P(N_1=k\,|\,N_1+N_2=n)=\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!}\cdot\frac{n!}{(\lambda_1+\lambda_2)^n}$$

$$=C_n^k(\frac{\lambda_1}{\lambda_1+\lambda_2})^k\cdot(\frac{\lambda_2}{\lambda_1+\lambda_2})^{n-k}\qquad(0\le k\le n)$$

$$\sim B(n,\frac{\lambda_1}{\lambda_1+\lambda_2})$$

即 N_1 在 $N_1+N_2=n$ 下的条件分布律与参数为 $(n,\dfrac{\lambda_1}{\lambda_1+\lambda_2})$ 的二项分布相同。

下面用两种方法求 $E(N_1\,|N_1+N_2=n)$

(i) 由定义式求：

$$E(N_1\,|\,N_1+N_2=n)=\sum_{k=0}^{n}k\cdot P(N_1=k\,|\,N_1+N_2=n)$$

$$=\sum_{k=o}^{n}k\cdot\frac{n!}{(\lambda_1+\lambda_2)^n}\cdot\frac{\lambda_1^k\lambda_2^{n-k}}{k!(n-k)!}=\frac{n\lambda_1}{(\lambda_1+\lambda_2)^n}\sum_{k=1}^{n}\frac{(n-1)!\lambda_1^{k-1}\lambda_2^{n-k}}{(k-1)!(n-k)!}$$

$$=\frac{n\lambda_1}{(\lambda_1+\lambda_2)^n}\cdot(\lambda_1+\lambda_2)^{n-1}=\frac{n\lambda_1}{(\lambda_1+\lambda_2)}$$

得到： $E(N_1\,|N_1+N_2)\ =\dfrac{\lambda_1}{(\lambda_1+\lambda_2)}(N_1+N_2)$ 。

（ii） $P(N_!=k\,|\,N_1+N_2=n)=C_n^k(\dfrac{\lambda_1}{\lambda\ \ +\lambda_2})^k\cdot(\dfrac{\lambda_2}{\lambda_1+\lambda_2})^{n-k}\sim B(n,\dfrac{\lambda_1}{\lambda_1+\lambda_2})$

故 $E(N_1\,|\,N_1+N_2)=\dfrac{\lambda_1}{(\lambda_1+\lambda_2)}\cdot(N_1+N_2)$

这说明 $E(N_1\,|\,N_1+N_2)$ 是 (N_1+N_2) 的线性函数。

由直观意义（ii），可得 $E(N_1|\,N_1+N_2)$ 的分布律：

$$P\{E(N_1\,|\,N_1+N_2)=\frac{n\lambda_1}{\lambda_1+\lambda_2}\}=P\{N_1+N_2=n\}=\frac{(\lambda_1+\lambda_2)^n}{n!}\cdot e^{-(\lambda_1+\lambda_2)}\qquad(n\in\mathrm{N})$$

（2）$E(N_1+N_2|N_1)$的分布律

$E(N_1+N_2|\,N_1)=E(N_1|\,N_1)+E(N_2|\,N_1)$ (性质 2)

$=N_1+E(N_2)$ (性质 4)

$=N_1+\lambda_2$

所以 $E(N_1+N_2|N_1=n)=n+\lambda_2 \quad (n\in N)$

$P\{E(N_1+N_2\mid N_1)=n+\lambda_2\}=P(N_1=n)=\dfrac{\lambda_1^n}{n!}\cdot e^{-\lambda} \quad (n\in \mathrm{N})$

例 5.3: 设：$(Y_n, n\geq 1)i.i.d$，$P(Y_n=1)=p\geq 0$，$P(Y_n=-1)=q=1-p\geq 0$ $X_0=0$，

$X_n=\sum_{k=1}^{n} Y_n$。求 $E(X_3|X_2)$的分布律，并证明 $E(X_{n+1}|X_n)=X_n+(p-q)$。

解:（1）$E(X_3|X_2)=E(X_2+Y_3|X_2)=E(X_2|X_2)+E(Y_3|X_2)=X_2+E(Y_3)=X_2+p-q$

$E(X_3|X_2)$的分布律如下表所示：

X_2	-2	0	2
$E(X_3\|X_2)$	$-2+(p-q)$	$p-q$	$2+p-q$
$P\{E(X_3\|X_2)\}=X_2+p-q$	$(1-p)^2$	$2p(1-p)$	p^2

（2）证明 $E(X_{n+1}|X_n)=X_n+(p-q)$。因为 $X_{n+1}=Y_{n+1}+X_n$，又 X_n 与 Y_{n+1} 独立，所以

$E(X_{n+1}|X_n)=E(Y_{n+1}+X_n|X_n)=E(X_n|X_n)+E(Y_{n+1}|X_n)=X_n+EY_{n+1}=X_n+(p-q)$

2 以连续型随机变量作为条件的条件数学期望

（1）定义

定义 5.4 对二维 $r.v.(X,Y)$，设 Y 为 $c.r.v.$，具有 $p.d.f.$为 $f_Y(s)$，$f_{X|Y=s}(t|s)$为 X 关于 $Y=s$ 的条件概率密度函数（简记 $c.p.d.f.$）；若 $\int_{-\infty}^{+\infty}|x|\, f_{X|Y=s}(x\mid s)dx<+\infty$，则称：

$$E(X\mid Y=s)\triangleq\int_{-\infty}^{+\infty} xf_{X|Y=s}(x\mid s)dx \qquad (5.3)$$

为 X 关于 $Y=s$ 的条件数学期望。

类似于第三章§4 可以定义 X 关于 $Y\in L$ 的条件分布函数及密度函数。事实上，对于给定的 $L\in B$,（即对一维波雷尔(*Borel*)集 L），若 $P(Y\in L)>0$，则 X 关于 $Y\in L$ 的条件分布函数定义为：$P(X\leq t\mid Y\in L)=\dfrac{P(X\leq t, Y\in L)}{P(Y\in L)}$；若对$\forall B\in\mathrm{B}$，存在非负函数 $f(t|L)$，满足：

$$P(X\in B\mid Y\in L)=\int_{u\in B} f(u\mid L)du$$

称 $f(t|L)$为 X 关于 $Y\in L$ 下的条件概率密度函数($c.p.d.f.$)。

例 5.4 设 $X\sim N(2,1^2)$,$L=\{X\geq 0\}$，求：X 在$\{X\geq 0\}$下的条件分布函数及条件 *c.p.d.f.*。

解： 当 $x<0$ 时，$P(X\leq x|X\geq 0)=0$,

当 $x\geq 0$ 时

$$P(X\leq x\mid X\geq 0)=\frac{P(0\leq X\leq x)}{P(X\geq 0)}=\frac{[\Phi(x-2)-\Phi(0-2)]}{1-\Phi(-2)}\text{。}$$

故
$$F_{X|X\geq 0}(x)=\begin{cases}0 & x<0,\\ (0.9772)^{-1}\cdot\dfrac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{(u-2)^2}{2}}du & x\geq 0,\end{cases}$$

又
$$F'_{X|X\geq 0}(x)=\begin{cases}0 & x<0,\\ (0.9772)^{-1}\cdot\dfrac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}} & x>0.\end{cases}$$

所以
$$f_{X|X\geq 0}(x)=\begin{cases}0 & x<0\\ (0.9772)^{-1}\cdot\dfrac{1}{\sqrt{2\pi}}e^{-\frac{(x-2)^2}{2}} & x\geq 0\end{cases}$$

对于一般的 *r.v.X*, $a<b$，求 X 在 $a<X<b$ 的条件分布（*c.d.f.*），通常称为 X 的截尾分布函数。

定义 5.5 给定 $L\in$ ✎，$P(Y\in L)>0$，若 $\int_{-\infty}^{+\infty}|t|\cdot f(t\mid L)dt<+\infty$，则称：

$$E(X\mid Y\in L)\hat{=}\int_{-\infty}^{+\infty}t\cdot f(t\mid L)dt \qquad (5.4)$$

为 X 关于 $Y\in L$ 下的**条件数学期望**。

如上例，$X\sim N(2,1^2)$，

$$E(X\mid X\geq 0)=\Phi^{-1}(2)\int_{-\infty}^{+\infty}e^{-\frac{(u-2)^2}{2}du}=\Phi^{-2}(2)(2+\Phi(2)-\frac{e^{-2}}{\sqrt{2\pi}})$$

显然，条件数学期望 $E(X|Y=y)$是 X 局限在$\omega\in\{Y=y\}$的局部加权平均，它是 y 的函数。类似于以 *d.r.v.*为条件的条件期望，定义一个新的随机变量 $E(X|Y)$，使其在$\omega\in\{Y=y\}$时的取值为 $E(X|Y=y)$。

定义 5.6 对二维 *r.v.*(X,Y)，设 Y 具有 *p.d.f.*为 $f_Y(y)$，对 $E|X|<+\infty$，$\forall y\in R$，$E(X|Y=y)$存在，若随机变量 $E(X|Y)$满足：

（1） $E(X|Y)$是 Y 的函数，且当$\omega\in\{Y=y\}$时，$E(X|Y)$的取值为：$E(X|Y=y)$;

（2） 对$\forall L\in$B，恒有：$E\{E(X|Y)|Y\in L\}=E\{X|Y\in L\}$。

则称随机变量 $E(X|Y)$为 X 关于 Y 的条件数学期望。

（2）广义的全概率公式

由上述定义，易知

$$E(X)=E\{E(X\mid Y)\}=\int_{-\infty}^{=\infty}E(X\mid Y=y)f_Y(y)dy=\int_{-\infty}^{+\infty}E(X\mid Y=y)dP(Y\le y) \qquad (5.5)$$

这是因为：注意到，当取 $L=R=(-\infty,+\infty)$时，

$$EX=E\{X|Y\in R\}=E\{E(X|Y)|Y\in(-\infty,+\infty)\}=E\{E(X|Y)\}=\int_{-\infty}^{+\infty}E(X\mid Y=y)f_Y(y)dy$$

当取 $X=I_A(\omega)$时，$E(X)=E(I_A)=P(A)$，又 $E(I_A|Y=y)=P(A|Y=y)$。则由(5.5)式有

$$P(A)=\int_{-\infty}^{+\infty}P(A\mid Y=y)f_Y(y)dy=\int_{-\infty}^{+\infty}P(A\mid Y=y)dP(Y\le y) \qquad (5.6)$$

(5.5)及(5.6)式可看作是广义的全概率公式。
应用(5.6)式，若取 $A=(X\le x)$, 则：

$$P(X\le x)=E(I_{(X\le x)}(\omega))=\int_{-\infty}^{+\infty}E(I_{(X\le x)}\mid Y=y)dP(Y\le y)$$

$$=\int_{-\infty}^{+\infty}P(X\le x\mid Y=y)dP(Y\le y)=\int_{-\infty}^{+\infty}P(X\le x\mid Y=y)dP(Y\le y)$$

得：

$$P(X\le x)=\int_{-\infty}^{+\infty}P(X\le x\mid Y=y)dP(Y\le y) \qquad (5.6a)$$

（3）条件数学期望的性质

设 EY，$E(X_i|Y)$，$E(h(Y))$，$E\{g(X)h(Y)\}$存在，则：

(*i*) $E[E(Y|X)]=E(Y)$；

(*ii*) $E(\sum_{i=1}^{n}C_iX_i\mid Y)=\sum_{i=1}^{n}C_iE(X_i\mid Y)$； *a.s.*

(*iii*) 若 X、Y 独立，$E[h(Y)|X]=E(h(Y))$； *a.s.*

(*iv*) $E(X|X)=X$，$E(g(X)|X)=g(X)$，*a.s.*

$$E[g(X)h(Y)|Y]=h(Y)E[g(X)|Y] \ a.s. \qquad (5.7a)$$

$$E[g(X)h(Y)]=E[h(Y)E[g(X)|Y]] \ a.s. \qquad (5.7b)$$

这些性质证明请读者思考并作为练习。

现就最后一个等式证明如下：
设(X,Y)有 *j.p.d.f.* $f(x,y)$，设 $f_Y(y)$是 Y 的 *p.d.f.*，于是

$$E\{g(X)h(Y)\}=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}g(x)h(y)f(x,y)dxdy==\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}[\frac{g(x)\cdot f(x,y)}{f_Y(y)}\cdot dx]h(y)f_Y(y)dy$$

$$=\int_{-\infty}^{+\infty}E[g(X)\mid Y=y]h(y)f_Y(y)dy=E[h(Y)E[g(X)|Y]]$$

特别地，若取 $I_A(\omega)=g(X)$，$h(Y)=1$ 得：$P(A)=\int_{-\infty}^{+\infty}P(A|Y=y)f_Y(y)dy$，这正是全概率公式!

下面举几个例子以加深对上述概念及性质的理解和认识。

例 5.5 设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，求 $E(Y|X)$。

解： 先求 Y 关于 $X=x$ 的 *c.p.d.f.*。

$$f_{Y|X=x}(y|x)=\frac{f(x,y)}{f_X(x)}$$

$$=\frac{1}{\sqrt{2\pi}\sigma_2(1-\rho^2)^{1/2}}\exp\{-\frac{1}{2\sigma_2^2(1-\rho^2)}[y-\mu_2-\rho\sigma_2\sigma_1^{-1}(x-\mu_{1)})]^2\}$$

即 $$f_{Y|X=x}(y|x)\sim N[\mu_2+\rho\sigma_2\sigma_1^{-1}(x-\mu_1),\sigma_2^2(1-\rho^2)]$$

$$E(Y|X=x)=\int_{-\infty}^{+\infty}y\cdot f_{Y|X=x}(y|x)dy=\mu_2+\rho\sigma_2\sigma_1^{-1}(x-\mu_1)$$

得： $E(Y|X)=\mu_2+\rho\sigma_2\sigma_1^{-1}(X-\mu_1)$。 (5.8)

从此例中我们看到：二元正态分布（X,Y）的条件数学期望是 X 的线性函数，这也正是正态分布的重要性质之一。

例 5.6 设 X_1 ,X_2 独立，$X_i\sim Ex(\lambda_i)$, $i=1,2$。求：$E(X_1|X_1\le X_2)$。

解：$P(X_1>x|X_1\le X_2)=\frac{P(X_1>x,X_1\le X_2)}{P(X_1\le X_2)}$

其中：$P(X_1>x, X_1\le X_2)=\int_x^\infty P(X_1>x,u\le X_2|X_1=u)\lambda_1e^{-\lambda_1u}du=\int_x^\infty P(u\le X_2)\lambda_1e^{-\lambda_1u}du$

$=\int_x^\infty\lambda_1e^{-(\lambda_1+\lambda_2)u}du=\frac{\lambda_1}{\lambda_1+\lambda_2}e^{-(\lambda_1+\lambda_2)x}$

$P(X_1\le X_2)=\int_0^{+\infty}P(X_2\ge x|X_1=x)dP(X_1\le x)=\int_0^{+\infty}e^{-\lambda_2x}\lambda_1e^{-\lambda_1x}dx=\frac{\lambda_1}{\lambda_1+\lambda_2}$

所以：$E(X_1|X_1\le X_2)=\int_0^{+\infty}P(X_1>x|X_1\le X_2)dx=\int_0^{+\infty}e^{-(\lambda_1+\lambda_2)x}dx=\frac{1}{\lambda_1+\lambda_2}$ 。

下面，我们来证明与这个结果相关的指数分布的一个有趣性质：

命题 5.1 设 X_1, X_2 独立，且 $X_i\sim Ex(\lambda_i)$,$i=1,2$, 则：

$$E(X_1|X_1\le X_2)=E[min(X_1,X_2)]$$

证明：因为：

$$P(min(X_1X_2)>t)=P(X_1>t,X_2>t)=e^{-(\lambda_1+\lambda_2)t}$$
$$P(min(X_1,X_2)\le t)=1-e^{-(\lambda_1+\lambda_2)t}$$

得 $min(X_1,X_2)\sim Ex(\lambda_1+\lambda_2)$

由例 5.6 的结论： $E(X_1|X_1 \le X_2)=E[min(X_1,X_2)]=\dfrac{1}{\lambda_1+\lambda_2}$

若推广这个结论可得：

命题 5.2 若 $X_1, X_2 \cdots\cdots ,X_n$ 独立，且 $X_i \sim E(\lambda_i), i=1, \cdots ,n$, 则：

$$min(X_1, \cdots, X_n) \sim E(\lambda_1+\lambda_2+\cdots+\lambda_n)$$

$$E(\min_{1\le i\le n} X_i)=(\sum_{i=1}^{n}\lambda_i)^{-1}\text{。}$$

例 5.7 设 X_1 ,X_2, X_3 独立，$X_i \sim Ex(\lambda_i)$, $i=1,2,3$。求：$E(X_2|X_1 \le X_2 \le X_3)$。

解：$E(X_2|X_1\le X_2\le X_3) = E(X_1+(X_2-X_1)|X_1\le X_2\le X_3)= E(X_1|X_1\le X_2\le X_3)+E(X_2-X_1| X_1\le X_2 \le X_3)$

由命题 5.2 知：$E(X_1|X_1\le X_2\le X_3)=\dfrac{1}{\lambda_1+\lambda_2+\lambda_3}$，根据指数分布的无记忆性，有：

$$E(X_2-X_1|X_1\le X_2\le X_3)= E(X_2| X_2\le X_3)= E(X_2\wedge X_3)=\frac{1}{\lambda_2+\lambda_3}\text{。}$$

故：$E(X_2|X_1 \le X_2\le X_3)=\dfrac{1}{\lambda_1+\lambda_2+\lambda_3}+\dfrac{1}{\lambda_2+\lambda_3}$。

3 条件概率、条件分布函数的推广

现在我们用条件数学期望的概念对条件概率、条件分布函数的概念加以推广。

设二维随机变量(X,Y)及任一随机事件 $B\in\mathrm{B}$，记 I_B 为 B 的示性函数。则 $P(B)=E(I_B(\omega))$。

我们称：$P(B|Y)\hat{=}E(I_B(\omega)|Y)$ 为事件 B 关于随机变量 Y 的**条件概率**；$\forall x\in R$，取 $B=(\omega:X\le x)$ 则称：$F(x|Y)\hat{=}P(X\le x|Y)=E(I_{(X\le x)}|Y)$为 X 关于 Y 的**条件分布函数**。

这样，前面我们介绍过的条件概率及条件分布函数都可以用条件数学期望来统一处理了。有些等式可表示的更为简练，例如：$\forall A\in$☞，二维 $r.v.(X,Y)$，$\forall x\in\mathrm{R}$，有 $P(A)=E(P(A|Y)), P(X\le x)=E[P(X\le x|Y)]$ 这正是(5.6)(5.6a)式的简洁形式。

4 多元随机变量的条件数学期望

定义 5.7 设三个 $d.r.v\ (X,Y,Z)$, 且 $E|X|<\infty$ ，则称：

$$E(X|Y=y_j,Z=z_k)=\sum_i x_i p(X=x_i|Y=y_{i,}Z=z_k)$$

为 X 关于 $Y=y_i, Z=z_k$ 的条件数学期望；称：

$$E(X|Y,Z)=\sum_j\sum_k I_{(Y=y_i),(Z=z_k)}\ E(X|Y=y_j,Z=z_k)$$

为 X 关于(Y, Z)的**条件数学期望**。

由定义，易知 $E(X|Y,Z)$具有以下两方面的直观意义：

（1）$E(X|Y,Z)$是(Y,Z)的二元函数，当 $Y=y_j,Z=z_k$ 时，$E(X|Y,Z)=E(X|Y=y_j,Z=z_k)$；

（2）对任意$(D_1,D_2)\in B^2$，$E[E(X|Y,Z)|Y\in D_1,Z\in D_2]=E[X|Y\in D_1,Z\in D_2]$。

注意：$E(X|Y,Z)$是(Y,Z)的二元函数，因此这样的条件数学期望刻划了 X 与(Y,Z)之间的概率特性。

对于 X 是一般 *r.v.*，而(Y,Z)是 *d.r.v* 或是 *c.r.v.*的情形，定义 $E(X|Y,Z)$与前面定义 $E(X|Y)$相类似。不难把以上定义推广到 n 元的情况，请读者自己尝试着定义。

以下是多元随机变量条件数学期望的一些简单性质。

若 $E|X|<\infty$，$E|X_i|<\infty$，$(1\le i\le n)$, $E|g(Y_1,\cdots,Y_n)|<\infty$，则有：

（1） $E(C_1X_1+\cdots+C_nX_n\,|\,Y_1,\cdots,Y_n)=\sum_{k=1}^{n}C_iE(X_i\,|\,Y_1,\cdots,Y_n)$； *a.s.*

（2）$E[g(Y_1,\cdots,Y_n)X|\,Y_1,\cdots,Y_n]=g(Y_1,\cdots,Y_n)E[X|\,Y_1,\cdots,Y_n]$； *a.s.*

（3）若 X 与 $Y_1,\cdots,Y_n$ 独立，则：$E[X|\,Y_1,\cdots,Y_n]=EX$； *a.s.*

（4）$E[E(X|\,Y_1,\cdots,Y_n)]=EX$，表明各局部加权平均的加权平均等于总的加权平均；

（5）对(X,Y,Z)的情况， 有：

$$E(X|Y)=E[E(X|Y,Z)|Y]=E[E(X|Y)|Y,Z]\ a.s. \tag{5.9}$$

性质（1）~（4）的证明均请读者自己补出。以下给出性质（5）在 *d.r.v.*情形下的证明：

先令 $Y=y_j$，$Z=z_k$，则有：

$E[E(X|Y)|\,Y=y_j,Z=z_k]=E[E(X|Y=y_j)|\,Y=y_j,Z=z_k]=E(X|Y=y_j)\,E[1|Y=y_j,Z=z_k]=E(X|Y=y_j)$；

故：$E(X|Y)=E[E(X|Y)|Y,Z]$；又因为

$$E[E(X\,|\,Y,Z)\,|\,Y=y_j]=E[E(X\,|\,Y,Z)\,|\,Y=y_j]$$

$$\begin{aligned}
&=\sum_k E(X\,|\,Y=y_j,Z=z_k)P(Z=z_k\,|\,Y=y_j)\\
&=\sum_k\sum_i x_iP(X=x_i\,|\,Y=y_j,Z=z_k)P(Z=z_k\,|\,Y=y_j)\\
&=\sum_k\sum_i x_iP(X=x_i,Z=z_k\,|\,Y=y_j)\\
&=\sum_i x_i\sum_k P(X=x_i,Z=z_k\,|\,Y=y_j)\\
&=\sum_i x_iP(X=x_i\,|\,Y=y_j)\\
&=E(X\,|\,Y=y_j)
\end{aligned}$$

即： $E(X|Y)=E[E(X|Y,Z)|Y]$

上式的直观意义留给读者自己解释。最好举一实际例子,以加深直观理解。

对于条件是连续型的随机向量的条件数学期望的定义，性质与以上相类似，不再赘述。

5 条件数学期望的应用：最佳均方预测

对二元随机变量(X,Y)，设 X 可观测，Y 不可观测，但 X 与 Y 之间有一定的统计关系。这时，希望用 X 的某一（*Borel* 可测）函数 $g(X)$作为 Y 的预测，记作$\tilde{Y}=g(X)$，并力求选取这样的$\tilde{Y}$ 使 $E(Y-\tilde{Y})^2$ 达到最小。若$\tilde{Y}^*$满足：$E(Y-\tilde{Y}^*)^2=\min[E(Y-g(X))^2]$，则称：$\tilde{Y}^*$为 Y 的**最佳均方预测**。

定理 5.1 若 $E(Y|X)$存在，则$\tilde{Y}^*=E(Y|X)$。

证明：等价于证明 $E[(Y-E(Y|X))^2]\le E[(Y-g(X))^2]$，对任意 *Borel* 可测函数 $g(X)$。

$$E[(Y-g(X))^2]=E\{(Y-E(Y|X))+(E(Y|X)-g(X))\}^2$$
$$=E[(Y-E(Y|X))^2]+E[(E(Y|X)-g(X))^2]+2E[(Y-E(Y|X))(E(Y|X)-g(X))]$$

而

$$E[(Y-E(Y|X))(E(Y|X)-g(X))]=E\{E[(Y-E(Y|X))(E(Y|X)-g(X))]|X\}$$
$$=E\{[E(Y|X)-g(X)]E[E(Y|X)-E(Y|X)]\}=0$$

得：$E[(Y-g(X))^2]=E[(Y-E(Y|X))^2]+E[(E(Y|X)-g(X))^2]\ge E[(Y-E(Y|X))^2]$

故：$\tilde{Y}^*=E(Y|X)$，即条件数学期望 $E(Y|X)$是对 Y 的最佳均方预测。

这表明用 $E(Y|X)$描述 Y 与 X 之间的关系是最佳的刻划（均方意义下）。它比用ρ_{XY}仅刻划两者之间线性关系的密切程度要更深入与精细。

在本定理的证明中，充分应用了条件数学期望的性质。可见，恰当地运用这些性质往往可以使问题得到大大的简化。

例 5.8 设$(X,Y)\sim N(\mu_1,\mu_2,\sigma_1^2,\sigma_2^2,\rho)$，用 X 预测 Y，求$\tilde{Y}^*$。

解：$\tilde{Y}^*=E(Y|X)$，由例 5.5，$E(Y|X)=\mu_1+\rho\sigma_2\sigma_1^{-1}(X-\mu_1)$，故

$$\tilde{Y}^*=E(Y|X)=\mu_2+\rho\sigma_2\sigma_1^{-1}(X-\mu_1) \tag{5.8}$$

进一步推广：

若用 $X_1,X_2,\cdots,X_n$ 的某一（*Borel* 可测）函数作为 X_{n+1} 的预测，则 X_{n+1} 的最佳均方预测为：$\tilde{X}_{n+1}^*=E(X_{n+1}|X_1\ \cdots,X_n)$，如右边存在的话。

例 5.9 设$\{\varepsilon_n,n\ge1\}$ *i.i.d*，且$\varepsilon_n\sim N(0,\sigma^2)$，$X_0=0$，$X_n-aX_{n-1}=\varepsilon_n$ ($|a|<1$ 常数)，用已知$X_1,X_2,\cdots,X_n$去预测 X_{n+1}，求 X_{n+1} 的最佳均方预测。

解：由 $X_k=\sum_{h=1}^{k-1}a^h\varepsilon_{k-h},1\le k\le n,$ 知 $X_1,X_2\cdots X_n$ 与ε_{n+1}独立，

$$\begin{aligned}\widetilde{X}_{n+1}^* &= E(X_{n+1} \mid X_1,\cdots,X_n) = E(aX_n + \varepsilon_{n+1} \mid X_1,\cdots X_n)\\ &= E(aX_n \mid X_1,\cdots.X_n) + E(\varepsilon_{n+1} \mid X_1,\cdots,X_n) = aX_n + E\varepsilon_{n+1}\\ &= aX_n\end{aligned}$$

6 条件方差

定义 5.8 设二维 $r.v.(X,Y)$，若 $E(X^2|Y)$存在，则称 $D(X \mid Y) = E[(X - E(X \mid Y))^2 \mid Y]$ 是 X 关于 Y 的**条件方差**。

易知：$D(X|Y)=E(X^2|Y)-(E(X|Y))^2$。

§6 母函数

对于取值非负整数的 $r.v.$，其母函数有极其良好的分析性质且又便于计算和分析，因此引入母函数是非常必要的。母函数又称生成函数(*Generating function*)。

定义 6.1 对于数列$\{a_n, n\geq 0\}$，称幂级数 $g(s)=\sum_{n=0}^{\infty} a_n s^n \ (|s|\leq 1)$ 为$\{a_n, n\geq 0\}$的母函数。

例 6.1 对组合数列 $C_n^0, C_n^1, \cdots, C_n^n$，其母函数为 $g(s)=\sum_{k=0}^{n} C_n^k s^k = (1+s)^n$ 。

定义 6.2 设 X 为非负整值随机变量，它的概率分布为 $P(X=k)=p_k$, k=0,1,2⋯，则称

$$g(s) \triangleq E(s^X)= \sum_{k=0}^{\infty} p_k s^k \quad |s|\leq 1 \tag{6.1}$$

为 X 的**概率母函数**，简称为**母函数**。

1 几个常用分布的母函数

（1）二项分布

设 $X\sim B(n,p)$，则它的母函数为：$g(s)=(q+sp)^n$

$$g(s)=E(s^X)= \sum_{k=0}^{n} p_k s^k = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} s^k = (q+sp)^n$$

由上式可得二点分布的母函数为：$g(s)=q+sp$。

（2）泊松分布

设 X 服从 $X\sim Po(\lambda)$，则其母函数为：$g(s)=e^{\lambda(s-1)}$

因

$$g(s)=E(s^X)=\sum_{k=0}^{\infty}\frac{\lambda^k e^{-\lambda}}{k!}s^k=e^{-\lambda}\cdot e^{s\lambda}=e^{\lambda(s-1)}$$

（3）几何分布

设 X 服从参数为 p 的几何分布，即 $P(X=k)=q^{k-1}p$, $k=1,2,\cdots$，则其母函数为：$g(s)=ps/(1-qs)$。

$$g(s)=E(s^X)=\sum_{k=1}^{\infty}q^{k-1}ps^k=\frac{ps}{1-qs}$$

2 母函数的基本性质

（1） $p_k=\dfrac{g^{(k)}(0)}{k!}$ $\qquad k\in N_0=\{0,1,2,\cdots\}$

若已知 X 的母函数，则由 $p_k=\dfrac{g^{(k)}(0)}{k!}$ 唯一地确定$\{p_k,k=0,1,\cdots\}$。所以说，X 的概率分布$\{p_k,k=0,1,\cdots\}$与其母函数 $g(s)$是一一对应的关系。

（2） 设非负整值随机变量 $X_1,X_2,\cdots X_n$ 相互独立, 而 $g_1,g_2,\cdots g_n$ 分别是它们的母函数，则 $Y=\sum_{k=1}^{n}X_k$ 的母函数为：

$$g(s)=g_1(s)g_2(s)\cdots g_n(s)=\prod_{k=1}^{n}g_k(s) \qquad (6.2)$$

证明：由定义可知：

$$g(s)=E(s^{X_1+\cdots+X_n})=E(s^{X_1}\cdot\cdots\cdot s^{X_n})=E(s^{X_1})\cdot\cdots\cdot E(s^{X_n})=\prod_{k=1}^{n}g_k(s)$$

上式第三个等号是由于 $X_1,\cdots X_n$ 独立。

该性质为寻找相互独立（取值非负整数的）随机变量之和的概率分布带来方便。见下面举例。

3 几个应用

例 6.2 设 $X_1,X_2\cdots X_n$ *i.i.d*， $X_k\sim B(1,p),1\le k\le n$，求 $Y=\sum_{k=1}^{n}X_k$ 的分布。

解：X_1 服从两点分布，其母函数为：$g_1(z)=q+pz$；又由于 X_i 同分布($i=1,2,\cdots n$)，由定理 6.1 可知 $Y=\sum X_k$ 的母函数为：$g(z)=(q+pz)^n$，则由 $p_k=\dfrac{1}{k!}g^{(k)}(0)$ 可求得 Y 的分布律。

例 6.3 设 X_1,X_2 独立，且 $X_i\sim B(n_i,p)$, $i=1,2$，证明：$X_1+X_2\sim B(n_1+n_2,p)$。

证明： 由 $g_{X_k}(s)=(q+sp)^n$ 及（6.2）式有

$$g_{X_1+X_2}(s)=(sp+q)^{n_1}\cdot(sp+q)^{n_2}=(sp+q)^{n_1+n_2}$$

所以 $X_1+X_2\sim B(n_1+n_2,p)$ 证毕。

例 6.4 设 $T_1,T_2\cdots T_n$ $i.i.d$ 且 $T_n\sim G(p)$,$0<p<1$，参数同为 p，求 $S_n=\sum_{k=1}^{n}T_k$ 的分布。

解：T_1 服从几何分布。则其母函数为 $g_1(s)=\dfrac{ps}{1-qs}$，由于 T_i 独立同分布，则 S_n 的母函数为 $g(s)=(\dfrac{ps}{1-qs})^n$，然后，我们将其展成幂级数，从而找出 S_n 的分布：

$$(\frac{ps}{1-qs})^n=p^ns^n(1-qs)^{-n}=p^ns^n\sum_{k=0}^{\infty}\frac{n(n+1)\cdots(n+k-1)}{k!}(qs)^k$$

$$=\sum_{k=0}^{\infty}C_{n+k-1}^{k}p^nq^ks^{n+k}=\sum_{k=0}^{\infty}C_{n+k-1}^{n-1}p^nq^{(n+k)-n}s^{n+k}=\sum_{k=n}^{\infty}C_{k-1}^{n-1}p^nq^{k-n}s^k$$

所以 S_n 的分布为：$P(S_n=k)=C_{n-1}^{k-1}p^nq^{k-n},k=n,n+1,\cdots$

例 6.5 设 X_1,X_2 相互独立，且 $X_k\sim Po(\lambda_k),k=1,2$。证：$X_1+X_2\sim Po(\lambda_1+\lambda_2)$

证明：由已知 $X_k\sim Po(\lambda_k)$，知 $g_k(s)=e^{\lambda_k(s-1)}$，$k=1,2$，且 X_1,X_2 互相独立。。则 X_1+X_2 的母函数 $g(s)=g_1(s)\cdot g_2(s)=e^{(\lambda_1+\lambda_2)(s-1)}$，所以 $X_1+X_2\sim Po(\lambda_1+\lambda_2)$ 证毕。

这与第三章的有关结论相同，但证明较简单。足见母函数的威力。

4 母函数与数字特征及期望的关系

（1）母函数与数字特征的关系

定理 6.1 母函数与期望和方差间的关系可如下表示

$$E(X)=g'(1)，D(X)=Var(X)=g''(1)+g'(1)-[g'(1)]^2 \quad (6.3)$$

证明：由定义 $g(s)=\sum_{k=0}^{\infty}p_ks^k$ 两边对 s 求导，有：

$$g'(s)=\sum_{k=0}^{\infty}kp_ks^{k-1}=\sum_{k=1}^{\infty}kp_ks^{k-1}$$

因此 $g'(1)=\sum_{k=1}^{\infty}kp_k=EX$ ，而 $g''(s)|_{s=1}=\sum_{k=2}^{\infty}k(k-1)p_k=E[X(X-1)]=E(X^2)-E(X)$

所以 $DX=EX^2-E(X)^2=g''(1)+g'(1)-[g'(1)]^2$ 证毕

（2）母函数的期望表示

我们由母函数的定义 $g_X(s)=\sum_{k=0}^{\infty}s^kP(X=k)=E(s^X)$ 可知，X 的母函数就是 s^X 的数学期望,即在母函数与期望之间有紧密的联系，使用母函数的期望表示可以简化母函数的运算、证明。这一点在证明定理 6.1 时已经看到了。

例 6.6 在一次核反应中，某个粒子可能分裂为 2 或 3 个粒子或不分裂。这三种可能性相应的概率分别为 p_2、p_3 与 p_1，新粒子的分裂性态是相同的(与原有粒子分裂性态相同)，并且分裂的情形彼此独立，求两次反应以后粒子总数的分布。

解：令 X_1=第一次反应后的粒子数，X_2=两次反应后的粒子数。

由题意 $E(s^{X_1})=g(s)=p_1s+p_2s^2+p_3s^3$，求 X_2 的母函数 $E(s^{X_2})$。由于两次反应后的总粒子数，是由一次反应后各个粒子分别独立的分裂(或不分裂)产生的粒子总和，故有 $X_2=Y_1+Y_2+\cdots+Y_{X_1}$，这里 Y_i= 第一次反应后第 i 个粒子分裂产生的粒子数，i=1,2,3，Y_i 与 X_1 同分布，故 $g_i(s)=E(s^{X_1})$；由定理 6.1，$E(s^{X_2})=E(s^{Y_1+\cdots+Y_{X_1}})$，由于 X_1 是随机变量，应用条件期望性质有：$E(s^{X_2})=E[E(z^{Y_1+\cdots+Y_{X_1}}|X_1)]$。因为：

$$E(s^{Y_1+\cdots+Y_{X_1}}\mid X_1=k)=E(s^{Y_1+\cdots+Y_k}\mid X_1=k)=E(s^{Y_1+\cdots+Y_k})=E(s^{Y_1})E(s^{Y_2})\cdots E(s^{Y_K})=[g(s)]^k$$

有：$E(s^{Y_1+\cdots+Y_{X_1}}\mid X_1)=[g(s)]^{X_1}$，得 $E(s^{X_2})=E[(g(s))^{X_1}]$，故将 $E(s^{X_1})=g(s)$ 的 s 换成 $g(s)$即为 X_2 的母函数 $E(s^{X_2})=g[g(s)]$。

有关母函数的详细内容可参阅费勒(*Feller*) 名著“概率论及其应用”第二卷。

练习题

4.1 对某一目标进行射击，直到击中 r 次为止，如果每次射击的命中率为 p，求需射击次数的均值和方差。

4.2 求超几何分布的均值和方差。

4.3 若 X 的密度函数是偶函数，且 $E(X^2)<\infty$，试证：$|X|$ 与 X 不相关，但它们不相互独立。

4.4 设轮船横向摇摆的随机振幅 X 的概率密度为：$f(x)=Axe^{-x^2/2\sigma^2}I_{(x>0)}$。求（1）$A$；（2）遇到大于其振幅均值的概率；（3）$X$ 的方差。

4.5 某袋中装有 N 张标号 1 至 N 的票券，按放回方式逐张抽取，问：

(1)到第一张抽出的票券再次被抽出时为止，抽取的期望数是多少？

(2)到第一次出现重复时为止，抽取的期望数是多少？

4.6 从[0，1]中随机的选 n 个点，分别求这 n 个值的极大值，极小值和极差的均值。

4.7 设 $X\sim N(\mu,\sigma^2)$，$Y\sim Po(\lambda), P(Y=k)=\lambda^k e^{-\lambda}/k!, k\in N_0$，$X$，$Y$ 独立，求

$Z=X/(Y+1)$ 的 $p.d.f.$ 及 EZ。

4.8 设 $\{X_i,1\le i\le n\}$ $i.i.d.,X_i\sim U(0,t),X_{(1)},X_{(2)},\cdots,X_{(n)}$ 为其顺序统计量，

$S_n=\sum_{i=1}^{n}X_i$，求：$n=3$ 时的 $EX_{(1)},EX_{(2)},EX_{(3)}$ 和 ES_3。

4.9 设 $\{X_i,1\le i\le n\}$ 独立，且 $X_i\sim Ex(\lambda_i)\quad i=1,\cdots,n$。求:

(1) $E（X_1|X_1\le X_2),$

(2) $E(X_1|X_1\le X_2\le X_3)$ $E(X_2|X_1\le X_2\le X_3),E(X_3|X_1\le X_2\le X_3)$，

(3) $E(X_1|X_1\le X_2\le\cdots\le X_n)$，

(4) $E(X_1\wedge X_2\wedge\cdots\wedge X_n)$。试分析以上结果，并从中找出有益的结论。

4.10 设 X 为取值非负整数 $r.v.$，证：$EX=\sum_{n=1}^{\infty}P(X\ge n)$；设 $r.v.Y\geqslant 0$，分布函数为 $F(s)$，证：$EY=\int_0^\infty[1-F(s)]ds$。

4.11 某箱中装有 n 张标号 1 至 n 的票券，按放回逐张抽取。问：（1）到第一张抽出的票券再次被抽出为止，抽取的期望数是多少？（2）到第一次出现重复为止，抽取的期望数是多少？（3）若只抽取 m 次，最大票券号码的期望数是多少？如是以无放回方

式抽取 m 次（$m \leqslant n$），回答同一问题。

4.12 设 $\{Y_n, n \geq 1\}i.i.d. P(Y_n = 1) = p > 0, P(Y_n = -1) = 1 - p = q > 0,\ X_0 = 0$，

$X_n = \sum_{i=1}^{n} Y_i,\ A_1 = \{Y_1 + Y_2 + Y_3 = -1\},\ A_2 = \{Y_2 + Y_3 + Y_4 + Y_5 = -2\},\ B_1 = \{X_4 = 2\}, B_2 =$

$\{Y_3 + Y_4 + Y_5 + \cdots + Y_9 = 3\}$ 。（1）当 $p = 1/3$ 时，A_1, A_2 是否独立？相容？证明你的结论；(2) 求 $P(B_1 \mid B_2), P(B_1 B_2), P(B_1 \cup B_2)$；（3）求 $P(X_n = k)(-n \leq k \leq n)$；（4）求：$E(X_4 \mid Y_1, Y_2), E(X_4 \mid X_2)$ 的分布律；（5）求 $E(I_{A_2} \mid I_{A_1})$ 分布律。

4.13 设水电公司在指定时间内限于设备能力，其发电量 X（万 kw）～$U[10, 23]$（均匀分布），用户用电量 Y（万 kw）～$U[10, 20]$。假设 X 与 Y 独立，水电公司每供应 1kw 电得到 0.32 元的利润，但空耗每 kw 电损失 0.14 元，而如用户用电量超过供电量时，公司需从别处补电，每 1kw 电反而赔 0.20 元。求在指定时间内，该公司获利 Z 的数学期望。

4.14 设 N_1, N_2 独立，$N_i \sim p(\lambda_i), i = 1,2$，（1）求 $E(N_1 \mid N_1 + N_2 = n)$；（2）求 $E(N_1 \mid N_1 + N_2)$ 及 $E(N_1 + N_2 \mid N_1)$ 的分布律。

4.15 设 $U=aX+b$，$V=cY+d$，$a \cdot c>0$。证：$\rho_{UV} = \rho_{XY}$。

4.16 设 $X_1, X_2, i.i.d., X_i \sim N(0,1^2)$，求 $E(X_1 \wedge X_2)$ 及 $E(X_1 \vee X_2)$。

4.17 $X = (X_1, X_2)^T \sim N(\mu, \Sigma)$，其中 $\mu = (\mu_1, \mu_2)^T = (1\ 2)^T$，$\Sigma = \begin{pmatrix} 1 & 4/5 \\ 4/5 & 1 \end{pmatrix}$ 令 $Y = \begin{pmatrix} 3 & 2 \\ 2 & 3 \end{pmatrix} X$，（1）求 $Y = (Y_1, Y_2)^T$ 的协方差矩阵，$E(Y_2 \mid Y_1)$ 及 $E(Y_1 + Y_2)$；

(2)求 $E(X_2 \mid X_1 = x)$ 及 $D(X_2 \mid X_1 = x) \equiv E\{[X_2 - E(X_2 \mid X_1 = x)]^2 \mid X_1 = x\}$；

（3）试问：$\{X_2 - E(X_2 | X_1)\}$ 与 X_1 是否独立？证明你的结论。

4.18 设 $\{X_i, 1 \leq i \leq n\}\ i.i.d.$。（1）当 $X_i > 0$ 时，证：

$E\left\{\frac{X_1}{X_1+\cdots+X_n}\right\}=1/n$；（2）证：$\operatorname{cov}\left(\sum_{i=1}^{n}X_i,\sum_{i=1}^{n+m}X_i\right)=nDX_1$；（3）试问（1）的结果能否推广？请具体给出并证明。

4.19 设 $P(A)=0.5,P(B)=0.6,P(B\mid A)=0.8$，求 $E(I_{A\cup B}\mid I_A=1),\quad E(I_{A\cup B}\mid I_A=0),$

及 $E(I_{A\cup B}\mid I_A),E(I_{A\cup B}\mid I_A,I_B)$ 的分布律。

4.20 设 A，B，$C\in\Phi$，试用定义证1（1）$E(E(I_A\mid I_B))=E(I_A)$），（2）$E(I_A\mid I_B)=$

$E(E(I_A\mid I_B,I_C)\mid I_B)=E(E(I_A\mid I_B)\mid I_B,I_C)$。

4.21 设 $\{Y_n,n\geq 1\}i.i.d.P(Y_n=1)=p>0,P(Y_n=0)=r\geq 0,P(Y_n=-1)=q>0,$

$p+q+r=1$,令 $X_0=0$ ，$X_n=\sum_{i=1}^{n}Y_i,A=\{Y_1+Y_2+Y_3=1\},\ B=\{X_4-X_1=2\}$。

（1）求 $P(AB),P(B|A),P(A\cup B)$；（2）求 $EX_{100},DX_{100},\operatorname{cov}(X_{100},X_{400})$ 及 X_{100} 与 X_{400} 的相关系数； （3）求 $E(I_B\mid I_A=1),E(I_B\mid I_A=0)$ 及 $E(I_B\mid I_A)$ 的分布律；（4） 求 $E(X_6\mid I_A=1)$ 及 $E(X_6\mid I_A)$ 的分布律。

4.22 设 $\{Y_n,n\geq 1\}i.i.d.P(Y_n=1)=p>0,P(Y_n=0)=r\geq 0,P(Y_n=-1)=q>0,$

$p+q+r=1$ 令 $X_0=0$，$X_n=\sum_{i=1}^{n}Y_i,U_n=X_n-n(p-q),V_n=\left(q/p\right)^{X_n}$。

（1）证 $EU_n=EU_0,EV_n=EV_0,\forall n\geq 2$；

(2)求 $E(U_3\mid U_0,U_1,U_2),E(V_4\mid V_0,V_1,V_2)$；

(3)证 $E(U_{n+1}\mid U_0,U_1,\cdots,U_n)=U_n,E(V_{n+1}\mid V_0,V_1,\cdots V_n)=V_n$；

(4)试考虑（3）的结论是否可以推广？

4.23 设 $\{Y_n,n\geq 1\}i.i.d.P(Y_n=1)=p>0,P(Y_n=0)=r\geq 0,P(Y_n=-1)=q>0,$

$p+q+r=1$，令 $X_0=0$，$X_n=\sum_{i=1}^{n}Y_i$。 记

$T=\min\{n:n\geq 1, X_n=-1或X_n=2\}$，试求 $P(X_T=-1)$ 及 $P(X_T=2)$。

4.24 设 $Y\sim U[0，1]$，令 $Y_n=\sum_{k=0}^{2^n-1}\frac{k}{2^n}I_{(\frac{k}{2^n}\leq Y<\frac{k+1}{2^n})}, n\geq 0$（1）求 $EY_2, E(YY_1), E(Y_2\mid Y_1)$；

（2）求 $EY_n, E(Y_3\mid Y_1,Y_2)$；（3）求 $E(Y_{n+1}\mid Y_n), E(Y_{n+1}\mid Y_1,\cdots,Y_n)$。

4.25 设 $\{Y_n, n\geq 1\}$ $i.i.d.$, $Y_k\sim N(0,\sigma^2), 1\leq k\leq n$，令 $X_n=\exp\left\{\frac{\mu}{\sigma^2}(Y_1+\cdots+Y_n)-\frac{n\mu^2}{2\sigma^2}\right\}$,

（1）求 $EX_1, E(X_2\mid X_1=x), \forall x\in R$；（2）求 $E(X_{n+1}\mid X_1,\cdots,X_n)$。

4.26 设 $Y\sim U[0,1]$，令 $Y_n=\sum_{k=0}^{2^n-1}\frac{k}{2^n}I_{(\frac{k}{2^n}\leq Y<\frac{k+1}{2^n})}, n\geq 0$, $X_n=\left(e^{\lambda(Y_n+2^{-n})}-e^{\lambda Y_n}\right)\cdot 2^n, n\geq 1$

（1）求 Y 在给定 $Y_1,\cdots,Y_n$ 下的条件 $p.d.f.$；（2）求 $E(X_2\mid X_1), E(X_3\mid X_1,X_2)$；

（3）求 $E(X_{n+1}\mid X_n), E(X_{n+1}\mid X_1,\cdots,X_n)$。

4.27 设 $\{X_i, 1\leq i\leq n\}$ $i.i.d.$，$X_n\sim U$（0,1），对 $\forall 0<t<1$ 记 $N_n(t)=\sum_{i=1}^n I_{(X_i\leq t)}$，

$\widetilde{F}_n(t)=\frac{1}{n}N_n(t), \alpha_n(t)=\sqrt{n}(\widetilde{F}_n(t)-t)$；（1）求 $EN_n(t), E\widetilde{F}_n(t)及D\widetilde{F}_n(t)$；

（2）$\forall 0<s<t\leq 1$，求给定 $N_n(s)$ 下 $N_n(t)-N_n(s)$ 的条件分布律；

（3）$\forall 0<s<t\leq 1$，求 $E(N_n(t)-N_n(s)\mid N_n(s))$ 的分布律；

（4）$\forall t\in[0,1]$，固定。证 $\forall\varepsilon>0, \lim\limits_{n\to\infty}P(|\widetilde{F}_n(t)-t|<\varepsilon)=1$。

4.28 设 $\{X_k, 1\leq k\leq n\}$ 独立且 $X_k\sim Ex(\lambda)$，$\lambda=\frac{1}{\theta}>0$，即 X_k 的分布函数为 $F(t)=(1-e^{-\lambda t})I_{(t\geq 0)}$，$X_{(1)},X_{(2)},\cdots,X_{(n)}$ 为其顺序统计量 $\forall t\geq 0, N_n(t)=\sum_{k=1}^n I_{(X_{(k)}\leq t)}$，

$\widetilde{F}_n(t)=N_n(t)/n$, $\widetilde{\theta}_{n,k}=\left(\sum_{i=1}^k X_{(i)}+(n-k)X_{(k)}\right)/k, 1\leq k\leq n$，（1）求证 $\forall t\geq 0$，$E(\widetilde{F}_n(t))=F(t)$；（2）记 $X_{(0)}=0$ 求 $(X_{(k)}-X_{(k-1)})$ 的 $p.d.f.$ 及 $E(X_{(k)}-X_{(k-1)})$，$1\leq k\leq n$；(3) 证 $X_{(1)},X_{(2)}-X_{(1)},\cdots,X_{(n)}-X_{(n-1)}$ 相互独立；(4) 求证 $E(\widetilde{\theta}_{n,k})=\theta=1/\lambda$。

4.29 已知 n 维随机向量 $X\sim N(0,I)$, $\Sigma=I$ 为 n 阶单位矩阵，试证：$X'\Sigma^{-1}X\sim\chi^2(n)$，

其中 $\chi^2(n)$ 的定义见题 3.30。

4.30 设 X～N（μ，Σ），其中 μ=0，$\Sigma^{-1}=\begin{pmatrix}6&3&2\\3&4&1\\2&1&2\end{pmatrix}$， 求 $(X_2,X_3)'$ 的联合密度以及 $X_1=x_1$ 的条件下，$(X_2,X_3)'$ 的条件密度。

4.31 设 N 同第二章 2.31 题中的定义，求 EN。

4.32 设 $\{X_k,1\le k\le n\}i.i.d.\ X_k\sim U[0,t]$，$X_{(1)},\cdots,X_{(n)}$ 为其顺序统计量，试求 $EX_{(k)},(1\le k\le n)$。

4.33 设 $r.v.X$，$E|X|<\infty$，$A\in\Phi$（A 为事件），$\{B_k,\ 1\le k\le n\}$ 为样本空间 Ω 的一个分解，$P(AB_k)>0,\ \forall k\in\{1,\cdots,n\}$，证：(1) $EX=\sum\limits_{k=1}^{n}P(B_k)E(X\mid B_k)$，

(2) $E(X\mid A)=\sum\limits_{k=1}^{n}P(B_k|A)E(X\mid AB_k)$。

4.34 设 $\widetilde{F}_n(x)$ 与 $F(x)$同第三章 3.33 题。(1) 求 $E\widetilde{F}_n(x)$，(2) 利用契贝雪夫不等式证明：$\forall\varepsilon>0,\lim\limits_{n\to\infty}P(|\widetilde{F}_n(x)-F(x)|<\varepsilon)=1$，(3) $\forall x\in R$ 固定，证明：$P(\lim\limits_{n\to\infty}\widetilde{F}_n(x)=F(x))=1$。

4.35 设(X,Y)，Z_1,Z_2 同第三章 3.34 题。(1) 求 $Cov(Z_1,Z_2)$，(2) 证明：Z_1 与 Z_2 相互独立。

4.36 设 $X=(X_1, X_2)^T\sim N(\mu,\Sigma)$为二维正态，令 $Y=AX$，其中 $A=(a_{ij})_{2\times2}$ 为非退化 2 阶矩阵，证明：$Y\sim N(A\mu,A\Sigma A^T)$。

4.37 求负二项分布 $P(X=k)=\binom{k+r-1}{r-1}p^r q^k, r\in N=\{1,2,\cdots\}\quad k=0,1,\cdots$ 的母函数。

4.38 设非负整值随机变量 X 的母函数为 $g(z)$，求 $X+1$ 及 $2X$ 的母函数。

4.39 求分布列，其母函数为：（1）$\frac{1}{4}(1+z)^2$；(2) $\frac{1}{2-z}$；(3) e^{z-1}。

4.40 直接找出下列特征函数的分布函数。(1) $\cos t$；(2) $\cos^2 t$。

4.41 设非负整值随机变量 X 的母函数为 g (z)，求：（1）$\sum\limits_{k=1}^{\infty}P(X\le k)z^k$；

（2）$\sum\limits_{k=0}^{\infty}P(X=2k)z^k$。

4.42 在重复独立实验中，每次实验成功的概率为 p，令 X 为首次在成功之后即遇到失败

的实验次数（即“X=n”意味着第 n–1 次实验成功，第 n 次实验失败，但在前 n–2 次实验中没有一次失败是紧跟在成功之后的）。求 X 的母函数，并找出 $E(X),Var(X)$。

4.43 设$\{Y_n, n\geq 1\}$是$i.i.d$非负$r.v.$，生成函数为$G_Y(s)$。$Z\sim G(1-p)$是参数为$(1-p)$的几何分布，即$P(Z=k)=(1-p)p^{k-1}, k\in N,$ 且Z与$\{Y_n, n\geq 1\}$独立，令$X=\sum_{n=1}^{Z}Y_n$，试求X的生成函数。

4.44 设X_1, X_2独立。（1）如$X_i\sim B(n_i, p), i=1,2$，证$X_1+X_2\sim B(n_1+n_2, p)$。

（2）如$X_i\sim P(\lambda_i), i=1,2$，证$X_1+X_2\sim P(\lambda_1+\lambda_2)$。

第五章 独立随机变量序列的极限定理

本章主要讨论随机变量序列一些基本的极限定理。包括：独立同分布随机变量序列的弱大数定律；*i.i.d. r.v.*序列的中心极限定理；特征函数及其性质；几种收敛概念与相互关系等；*i.i.d. r.v.*序列的强大数定律。

§1. 大数定律

这里研究的大数定律，主要内容是研究独立随机变量序列的前 n 项算术平均在什么条件下以概率收敛到它们的数学期望的平均。

众所周知，在 n 次独立重复 *Bernouli* 试验中，事件 A 出现的频率 $f_n(A)=\frac{\mu_n(A)}{n}$ 在一定概率意义下趋向事件 A 的概率，这早已为大量试验结果所证实。那么它的理论依据是什么？如何来确切刻划以上事实？下面定理给出答案。

定理 1.1 (*Bernouli* 大数定律) 设$\mu_n(A)$是 n 次 *Bernouli* 试验中事件 A 出现的次数，且 $P(A)=p$。则：$\forall\varepsilon>0$，

$$\mathop{Lim}_{n\to\infty} P\{\omega:|\frac{\mu_n}{n}-p|<\varepsilon\}=1\text{。}\tag{1.1}$$

证明： 令：$X_i=\begin{cases}1 & 第i次试验出现A\\ 0 & 第i次试验不出现A\end{cases}$， 则$\{X_i,\ 1\le i\le n\}$ *i.i.d.*，$X_i\sim B(1,p)$。$E\,X_i=p$，$D\,X_i=p(1-p)$。 $\mu_n=\sum\limits_{i=1}^n X_i$， $E\frac{\mu_n}{n}=E(\frac{1}{n}\sum\limits_{i=1}^n X_i)=p$， $D\frac{\mu_n}{n}=D(\frac{1}{n}\sum\limits_{i=1}^n X_i)=\frac{1}{n^2}D(\sum\limits_{i=1}^n X_i)=\frac{p(1-p)}{n}\le\frac{1}{4n}$，故由 *Chebyshev* 不等式知：$\forall\varepsilon>0$，

$$0\le P\{\omega:|\frac{\mu_n}{n}-p|\ge\varepsilon\}\le\frac{1}{\varepsilon^2}D(\frac{\mu_n}{n})\le\frac{1}{4n\varepsilon^2}\to 0\quad(n\to\infty)$$

得：
$$\mathop{Lim}_{n\to\infty} P\{\omega:|\frac{\mu_n}{n}-p|<\varepsilon\}=1\qquad\forall\varepsilon>0\text{。}$$

以上事实，我们称事件 A 的频率 $\frac{\mu_n}{n}$ 以概率收敛到 $p=P(A)$，简记为：

$\frac{\mu_n}{n} \xrightarrow{p} P(A)\ (n \to \infty)$。直观解释，由于 n□∞时，$D(\frac{\mu_n}{n}) = D(\frac{\mu_n}{n} - p) \to 0$，因而不管任给多么小的$\varepsilon$ >0，当 n 越大时，事件$\{\omega : |\frac{\mu_n}{n} - p| < \varepsilon\}$的概率就越大，且越接近于 1，即频率 $\frac{\mu_n}{n}$ 的取值落在 p 的ε小邻域中的概率越接近于 1，即 n□∞时，有 $P\{\omega : |\frac{\mu_n}{n} - p| < \varepsilon\} \to 1$。

对于一般 *i.i.d.*随机变量序列有以下结果：

定理 1.2 设$\{Y_n, n\geq 1\}$ *i.i.d*，且 $EY_n=\mu$，$DY_n=\sigma^2<\infty$。令 $X_n = \frac{1}{n}\sum_{k=1}^{n} Y_k$，则 $\forall \varepsilon > 0$,

$$\lim_{n\to\infty} P\{\omega : |X_n - \mu| < \varepsilon) = 1 \tag{1.2}$$

称$\{X_n, n\geq 1\}$以概率收敛到μ。简记为：$X_n = \frac{1}{n}\sum_{k=1}^{n} Y_k \xrightarrow{P} \mu \quad (n \to \infty)$。

证明：显然：$EX_n = \mu, DX_n = \frac{\sigma^2}{n}$。由 *Chebyshev* 不等式有：$\forall \varepsilon > 0$

$$0 \leq P\{\omega : |X_n - \mu| \geq \varepsilon\} \leq \frac{1}{\varepsilon^2} \cdot \frac{\sigma^2}{n}$$

故

$$1 - \frac{1}{\varepsilon^2}\frac{\sigma^2}{n} \leq P\{\omega : |X_n - \mu| < \varepsilon\} \leq 1$$

即 $\forall \varepsilon > 0$, $$\lim_{n\to\infty} P\{\omega : |X_n - \mu| < \varepsilon) = 1.$$

直观解释与应用：

（1）若μ表测量对象的真值（理论值）。Y_k表第 k 次测量值，X_n表前 n 次独立重复测量的平均值。显然它的均方误差 $E(X_n - \mu)^2 = \frac{1}{n}\sigma^2$ 要比 $E(Y_k - \mu)^2 = \sigma^2$ 小。若 ε 表测量误差限，(1.2)式表明不管精度要求多么高（即 ε 多么小），只要取测量次数 n 足够多，则取其平均值 X_n 就可使落在真值的误差限内的概率就可充分接近于 1。

（2）记 $A_n(\varepsilon) = \{\omega : |X_n - \mu| < \varepsilon\}$，(1.2)式等价于：$\forall \varepsilon > 0, P(A_n(\varepsilon)) \to 1\ (n \to \infty)$。

（3）注意此处 $\underset{n\to\infty}{Lim}\, P(A_n(\varepsilon)) \neq P(\underset{n\to\infty}{Lim}\, A_n(\varepsilon))$，而且此时 $\underset{n\to\infty}{Lim}\, A_n(\varepsilon)$ 不一定有意义！

例 1.1：（经验分布以概率收敛到理论分布）

设 $r.v.X \sim F(x)=P(X\le x)$，$\{X_k, 1\le k\le n\}$ $i.i.d.$ 且 X_k 与 X 同分布，$\forall x\in R$，称 $\widetilde{F}_n(x)=\sum_{k=1}^{n} I_{(X_k\le x)}/n$ 为 $r.v.X$ 的经验分布，则有：$\forall x\in R$，$n\ge 1$

$$(1)\quad E\widetilde{F}_n(x)=F(x),\quad D\widetilde{F}_n(x)=F(x)(1-F(x))/n \tag{1.3}$$

$$(2)\quad \forall \varepsilon>0,\quad \underset{n\to\infty}{Lim}\, P(|\widetilde{F}_n(x)-F(x)|<\varepsilon)=1. \tag{1.4}$$

简记：$\underset{n\to\infty}{Lim}\, \widetilde{F}_n(x)\overset{p}{=}F(x)$。

解：(1.3)显然，(1.4)式由(1.3)式再利用定理 1.2 即得。

应用：在实际工作中，有时理论分布 $F(x)$未知，可用 $\widetilde{F}_n(x)$ 作为 $F(x)$的估计。

上述定理可推广为以下的 *Markov* 大数定律。

定理 1.3 设$\{Y_n, n\ge 1\}$为 $r.v.s$，若 $\underset{n\to\infty}{Lim}\, \frac{1}{n^2}\cdot D(\sum_{k=1}^{n} Y_k)=0$，则 $\forall \varepsilon>0$

$$\underset{n\to\infty}{Lim}\, P\{\omega: |\frac{1}{n\cdot}\sum_{k=1}^{n} Y_k - \frac{1}{n}\cdot\sum_{k=1}^{n} EY_k| < \varepsilon\} = 1$$

证明：利用 *Chebyshev* 不等式即得。

§2 特征函数

第四章的讨论表明，数学期望、方差等数字特征只能粗略反映分布函数的某些性质，一般情况下并不能完全确定分布函数。本节引入特征函数的概念，既可完全刻画分布函数又可利用其良好的分析性质来解决某些问题。

在上一章中所研究的母函数为非负整值随机变量的研究带来了很大方便，特征函数则是针对一般的随机变量而设立的。

1 特征函数的定义

定义 2.1 如果 X,Y 均为概率空间(Ω,F,P) 上的实值 *r.v.*,，则称 $\xi=X+iY$ 为一复随机变量。且定义复 *r.v.*$\xi=X+iY$ 的数学期望为：$E\xi=EX+iEY$，其中 $i=\sqrt{-1}$ 为虚数。

则由以上定义知：$E(e^{itX})=E(\cos tX+i\sin tX)=E(\cos tX)+iE(\sin tX)$

定义 2.2：若 *r.v.*X 的分布函数为 $F_X(x)$，则称：

$$\varphi(t)\hat{=}Ee^{itX}=\int_{-\infty}^{+\infty}e^{itx}dF_X(x)\hat{=}\int_{-\infty}^{+\infty}(\cos tx+i\sin tx)dF_X(x)$$

为 X 的特征函数。(简记为 *c.f.*).

若 $g(X)$是一个一维 *Borel* 可测函数，$Y=g(X)$，则有：

$$Ee^{itY}=Ee^{itg(X)}=\int_{-\infty}^{+\infty}e^{itg(x)}dF_X(x).$$

其中 $e^{itY}=\cos tY+i\sin tY$.

简单性质：

（1）由于$|e^{itX}|\le 1$，故对任一随机变量 X 及实数 t 都有意义，即任一随机变量都具有 *c.f.*。

（2）*c.f.*是一个实变量的复值函数。

（3） *c.f.*只与分布函数有关，因此又称为某一分布函数的特征函数。

（4）若 X 的特征函数为$\varphi(t)$，则 $a+bX$ 的特征函数为：

$$\varphi_{a+bX}(t)=e^{ita}\varphi(bt)$$

（5）$\varphi(0)=Ee^0=1$。

（6）对 *d.r.v.*X，$P(X=x_j)=p_j, j=1,2,\cdots$，则 X 的特征函数为：

$$\varphi(t)=\sum_{j=1}^{\infty}p_je^{itx_j}.$$

对 *c.r.v.*X，其 *p.d.f.*为 $f(x)$，则 X 的特征函数为：

$$\varphi(t)=\int_{-\infty}^{+\infty}e^{itx}f(x)dx$$

2　几种常见分布的特征函数

（1）二点分布：

设 $X\sim B(1,p)$，$q=1-p$，则由定义可知：$\varphi(t)=e^{it}p+e^0q=e^{it}p+q$

（2）泊松分布：

设 $X\sim Po(\lambda)$，即：$P(X=k)=\dfrac{\lambda^k}{k!}e^{-\lambda}, k\in N$；则：$\varphi(t)=\sum_{k=0}^{\infty}e^{itk}\dfrac{\lambda^k}{k!}e^{-\lambda}=e^{-\lambda}e^{i\lambda t}=e^{\lambda(it-1)}$

（3）正态分布：

(i) 标准正态分布的 $c.f.$。设 X~$N(0,1^2)$，$\because \int_{-\infty}^{+\infty}\sin tx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0.$

$$\therefore \varphi(t)=\int_{-\infty}^{+\infty}e^{itx}\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\int_{-\infty}^{+\infty}\cos tx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=0.$$

两边对 t 求导，有：

$$\varphi'(t)=-\int_{-\infty}^{+\infty}x\sin tx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=\int_{-\infty}^{+\infty}\sin tx\frac{1}{\sqrt{2\pi}}\left(e^{-\frac{x^2}{2}}\right)'dx$$

$$=-\int_{-\infty}^{+\infty}t\cos tx\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}dx=-t\varphi(t)$$

即可得到微分方程：$\varphi'(t)+t\varphi(t)=0$, 加上$\varphi(0)=1$，有：$\varphi(t)=e^{-\frac{t^2}{2}}$

(ii) 对于一般的正态分布。设 X~$N(\mu,\sigma^2)$，可以这样来看：$Y=\frac{X-\mu}{\sigma}\sim N(0,1)$，

所以 $X=\mu+\sigma Y$；由简单性质（4）知：$\varphi_X(t)=e^{i\mu t}\varphi_Y(\sigma t)=e^{i\mu t-\frac{1}{2}\sigma^2t^2}$。

3 特征函数的性质

与母函数相似，对于一般随机变量都有定义的特征函数也具有很好的分析特性。它作为一种从分布函数到某类实变复值函数的变换不仅具有一一对应性，而且还有某种连续性。此外，它与数字特征也有直接的联系。

以下均设$\varphi(t)$为 $r.v.X$ 的特征函数。

性质 1 特征函数$\varphi(t)$，则 $|\varphi(t)|\le \varphi(0)=1$，且$\varphi(-t)=\overline{\varphi(t)}$。

证明：显然：$\varphi(0)=1$，而

$$|\varphi(t)|=|\int_{-\infty}^{\infty}e^{itx}dF(x)|\le\int_{-\infty}^{\infty}|e^{itx}|dF(x)\le\int_{-\infty}^{\infty}dF(x)=1=\varphi(0)$$

$$\varphi(-t)=\int_{-\infty}^{\infty}e^{i(-t)x}dF(x)=\int_{-\infty}^{\infty}e^{-itx}dF(x)=\overline{\varphi(t)}$$

性质 2 $\varphi(t)$ 在$(-\infty,+\infty)$上一致连续。

性质 3 若 $E(X^k)$存在，则对任 $t\in(-\infty,+\infty)$，$\varphi(t)$ k 阶可导，且

$$\varphi^{(k)}(t)=i^kE(X^ke^{itX}).\ \text{特别有：}\ \varphi^{(k)}(0)=i^kE(X^k).$$

性质 4 $\varphi(t)$具有非负定性，即对任意的正整数 n 及任意实数 $t_1, t_2,\cdots, t_n$ 与复数λ_1, λ_2

$\cdots\lambda_n$ 总有：$\sum_{k=1}^{n}\sum_{j=1}^{n}\varphi(t_k - t_j)\lambda_k\overline{\lambda_j} \geq 0$。

证明：显然对 $\forall t_k\in R$，λ_k 为复数，$1\leq k\leq n$，有：

$$|\sum_{k=1}^{n}\lambda_k e^{it_kX}|^2 = (\sum_{k=1}^{n}\lambda_k e^{it_kX})\overline{(\sum_{j=1}^{n}\lambda_j e^{it_jX})} \geq 0$$

故：

$$E|\sum_{k=1}^{n}\lambda_k e^{it_kX}|^2 = \sum_{k=1}^{n}\sum_{j=1}^{n}E(e^{i(t_k-t_j)X})\lambda_k\overline{\lambda}_j = \sum_{k=1}^{n}\sum_{j=1}^{n}\varphi(t_k - t_j)\lambda_k\overline{\lambda_j} \geq 0$$。

性质 5　设 $X_1,\cdots, X_n$ 相互独立，$\varphi_1,\cdots, \varphi_n$ 分别为它们的特征函数，则 $Y= X_1+\cdots+ X_n$ 的特征函数为 $\varphi(t)= \varphi_1(t)\,\varphi_2(t)\cdots\varphi_n(t)$。

证明：$\varphi(t) = Ee^{itY} = Ee^{it(X_1+\cdots+X_n)} = \prod_{k=1}^{n}Ee^{itX_k} = \prod_{k=1}^{n}\varphi_k(t)$。（第三个等号是根据 $X_1,\cdots, X_n$ 相互独立）。

性质 5 是特征函数的一条重要性质，独立性问题在概率论中具有重要的地位，而应用特征函数来处理独立问题是简洁明了的。因此 *c.f.* 在概率论中有其独特的作用。

4　特征函数与分布函数的关系

定理 2.1　分布函数 $F(x)$到特征函数$\varphi(t)$的变换是一一对应的。

注意　$\varphi(t) = Ee^{itX} = \int_{-\infty}^{+\infty}e^{itx}dF_X(x)$　因此，$\varphi(t)$实质上是由 X 的分布函数 $F(x)$决定的；所谓一一对应，这里指：若 $F_1(x)\neq F_2(x)$，则$\varphi_1(t)\neq\varphi_2(t)$。从$\varphi(t)$ 到 $F(x)$的具体公式如下：

$$F(b) - F(a) = \frac{1}{2\pi}\underset{T\to\infty}{Lim}\int_{-T}^{T}\frac{e^{-ita} - e^{-itb}}{it}\varphi(t)dt$$

其中 $a<b$，皆是 $F(x)$的连续点。

证明：略。有兴趣的读者可参看[3]，[5]。

定理 2.2　分布函数序列$\{F_n(x),n\geq1\}$与分布函数 $F(x)$具有关系：

$$\underset{n\to\infty}{Lim}F_n(x) = F(x)\quad (对任意\ F(x)的连续点)$$

当且仅当：$\underset{n\to\infty}{Lim}\varphi_n(t) = \varphi(t)$，其中 $f(t)$是 $F(x)$的特征函数,而 $f_n(t)$是 $F_n(x)$的 *c.f.*。

证明：略。有兴趣的读者可参看[3]，[5]。

定理 2.3 设$\{Y_n,n\geq1\}$ *i.i.d.*，且 $EY_n= \mu<\infty$，则对 $\forall\varepsilon > 0$,

$$\lim_{n\to\infty} P\{\omega:|\frac{1}{n}\sum_{k=1}^{n}Y_k-\mu|<\varepsilon)=1.$$

证明：由 $Y_1,Y_2,,\cdots,Y_n$ 有相同的分布，设它们的特征函数为φ(t)，由他们期望存在 $EY_n=\mu$.

故：φ(t)= φ(0)+ φ(0)t+o(t)=1+$i\mu t$+o(t)；从而 $\frac{1}{n}\sum_{k=1}^{n}Y_k$ 的特征函数$\varphi_n(t)$为

$$\varphi_n(t)=[\varphi(\frac{t}{n})]^n=[1+i\mu\frac{t}{n}+o(\frac{t}{n})]^n$$

对固定的 $t\in R$ 有 $\varphi_n(t)\to e^{i\cdot\mu\cdot t}(n\to\infty)$

而极限函数 $e^{i\cdot\mu\cdot t}$ 是连续函数，它是退化分布 $I_{(Y=\mu)}$对应的特征函数，由此可知：

$\frac{1}{n}\sum_{k=1}^{n}Y_k$ 的分布函数弱收敛于 $I_{(Y=\mu)}$ 由引理可知：$\frac{1}{n}\sum_{k=1}^{n}Y_k\xrightarrow{P}\mu$ 。

5 多元特征函数

设随机向量 X=(X_1,---, X_n)的 $j.p.d.f.$为 p(x_1,---, x_n)，类似于一维我们可定义随机向量 X 的特征函数为：

$$\varphi(t_1,t_2\cdots t_n)=E[\exp(i\sum_{k=1}^{n}t_kX_k)]=\int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\exp(i\sum_{k=1}^{n}t_kx_k)\cdot p(x_1,x_2\cdots x_n)dx_1dx_2\cdots dx_n$$

例：设 $X=(X_k,1\le k\le n)^T\sim N(\mu,\Sigma)$为 n 维正态 $r.v.$，则其特征函数为：

$$\varphi(t)=Ee^{it^TX}=\exp(i\mu^Tt-\frac{1}{2}t^T\Sigma t)\text{。}$$

对于 n 元特征函数的性质与一元相类似,读者可参阅复旦大学编著的第一册“概率论基础”及 *Feller* 的“概率论及其应用”第二卷。

§3 中心极限定理

在实际问题中，往往要考虑很多个随机因素对某一随机现象的指标所产生的总的影响，而这些指标的统计规律性常常呈现出近似于正态分布。这些事实就其原因是什么？

另一方面，对一串随机变量序列{Y_n, $n\geq1$}，对于其前 n 项和 $X_n=\sum_{k=1}^{n}Y_k$ 当 n 很大时求 X_n

的分布是件复杂的事情，自然希望（它们）在一定条件下 X_n 的极限分布能够存在，这是很有意义的工作。本节要介绍的中心极限定理就是研究 X_n 在什么条件下，它们的极限分布恰是正态分布。为使问题的提法更具一般化和应用性，这一节我们感兴趣的问题是：随机变量序列$\{Y_n,\ n\geqslant 1\}$的前 n 项和 X_n 经标准化后，即 $Z_n=\dfrac{X_n-EX_n}{\sqrt{DX_n}}$ 在什么条件下，它们的极限分布是标准正态分布。

1　独立同分布的情形

定理 3.1（***Lindeberg-levy* 定理**）设$\{Y_n, n\ge 1\}$ *i.i.d.*，且 $EY_1=\mu$，$DY_1=\sigma^2<\infty$，

记 $X_n=\sum_{k=1}^{n}Y_k$ ，$Z_n=\dfrac{X_n-EX_n}{\sqrt{DX_n}}=\dfrac{\sum_{k=1}^{n}(Y_k-\mu)}{\sqrt{n}\cdot\sigma}=\sum_{k=1}^{n}(\dfrac{Y_n-\mu}{\sqrt{n}\cdot\sigma})$，则$\forall x\in R$ 有

$$\underset{n\to\infty}{Lim}P(Z_n\le x)=\Phi(x)=\frac{1}{\sqrt{2\cdot\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du$$

简记：$P(Z_n\le x)\xrightarrow{w}\Phi(x)$

证明:令 $Y_n'=\dfrac{Y_n-\mu}{\sigma}$,则$EY_n'=0, DY_n'=1$. 设 Y_n' 的特征函数为 $\varphi(t)=Ee^{i\cdot t\cdot Y_n'}$,则

$$\begin{aligned}\varphi(t)&=1+i\cdot\varphi'(0)t+\frac{\varphi''(0)}{2!}t^2+o(t^2)\\&=1-\frac{t^2}{2}+o(t^2)\end{aligned}$$

$\dfrac{Y_n'}{\sqrt{n}}$ 的特征函数为:$\varphi(\dfrac{t}{\sqrt{n}})=1-\dfrac{t^2}{2\cdot n}+\dfrac{o(t^2)}{n}$。故 Z_n 的特征函数为

$$\varphi_n(t)=(1-\frac{t^2}{2\cdot n}+\frac{o(t^2)}{n})^n=\{[1-\frac{t^2}{2\cdot n}+\frac{o(t^2)}{n}]^{-\frac{2n}{t^2}}\}^{-\frac{t^2}{2}}$$

故:$\underset{n\to\infty}{Lim}\varphi_n(t)=e^{-\frac{t^2}{2}}$。$\forall x\in \mathrm{R}$

$$\underset{n\to\infty}{Lim}P(Z_n\le x)=\Phi(x)=\frac{1}{\sqrt{2\cdot\pi}}\int_{-\infty}^{x}e^{-\frac{u^2}{2}}du$$

以下是中心极限定理几个简单应用的例子：

（1）正态随机数的产生

计算机上通常可以生成服从[0,1]上均匀分布的“伪随机数”。如何由此产生正态分布的随机数呢？

设$\{Y_k，k\geq1\}$ *i.i.d.*，且 $Y_k \sim U[0,1]$，令 $X_n=\sum_{k=1}^{n}Y_k$ 。记 $f_n(x)$为 X_n 的 *p.d.f.*，则：

$$f_1(x)=\begin{cases}1 & x\in[0,1]\\ 0 & x\notin[0,1]\end{cases};$$

$$f_2(x)=\begin{cases}x & x\in[0,1]\\ 2-x & x\in[1,2]\\ 0 & \text{其它}\end{cases};$$

$$f_3(x)=\begin{cases}\frac{1}{2}x^2 & x\in[0,1]\\ \frac{1}{2}(-2x^2+6x-3) & x\in[1,2]\\ \frac{1}{2}(3-x)^2 & x\in[2,3]\\ 0 & \text{其它}\end{cases}$$

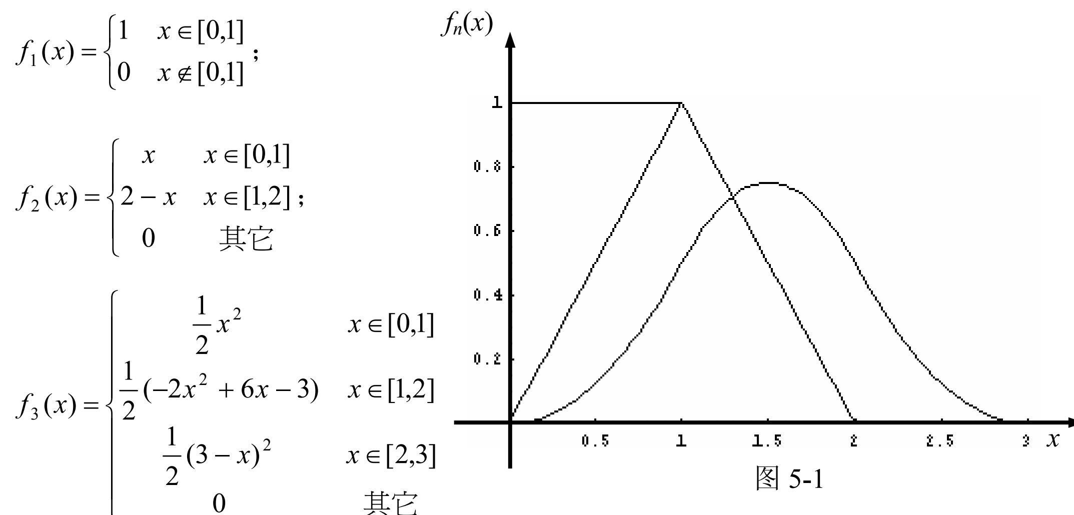


图 5-1

由图 5-1 可以看出，$f_3(x)$已接近于正态的情形，为将其标准化，令: $Z_n=\frac{X_n-EX_n}{\sqrt{DX_n}}$ 。

由中心极限定理有: $\forall x\in R$,

$$P(Z_n\leq x)\xrightarrow{w}\Phi(x)\quad(n\to\infty)$$

通常取 $Z_{12}=\sum_{k=1}^{12}Y_k-6$ 作为正态分布随机变量的近似。

（2）对误差限可靠度的估计与实验次数的确定。

在贝努利独立重复试验中，若事件 A 的概率 $P(A)=p$ 未知，可用 n 次试验中出现 A 的频率 $\frac{\mu_n(A)}{n}$ 作为 A 的估计。记 $\hat{p}=\frac{\mu_n(A)}{n}$ 。设估计的误差限为$\delta>0$，误差限的可靠度β定义为

$$\beta\triangleq P\{\omega:|\frac{\mu_n}{n}-p|<\delta\}$$

a) 给定δ及试验次数 n，以及试验结果，估计β。

由中心极限定理， 当 n 充分大时

$$\beta = P\{\omega : |\frac{\mu_n}{n} - p| < \delta\}$$
$$= P\left\{\omega : -\delta\sqrt{\frac{n}{p(1-p)}} < \frac{\mu_n(A) - np}{\sqrt{np(1-p)}} < \delta\sqrt{\frac{n}{p(1-p)}}\right\}$$
$$\doteq 2\Phi\left[\delta\sqrt{\frac{n}{p(1-p)}}\right] - 1$$

再以 $\widetilde{p} = \frac{\mu_n(A)}{n}$ 作为 p 的近似代入，则β 的估计 $\widetilde{\beta}$ 取为

$$\widetilde{\beta} = 2\Phi\left[\delta \cdot n \cdot \sqrt{\frac{n}{\mu_n(A)(n - \mu_n(A))}}\right] - 1$$

b)　给定δ 及β，如可确定试验次数 n 以满足要求。

即要求 n 满足

$$P\{\omega : |\frac{\mu_n(A)}{n} - p| < \delta\} \geq \beta$$

为此需　$2\Phi\left[\delta\sqrt{\frac{n}{p(1-p)}}\right] - 1 \geq \beta$　即：$\Phi\left[\delta\sqrt{\frac{n}{p(1-p)}}\right] \geq \frac{1+\beta}{2}$

记 Z_β满足 $\Phi(Z_\beta) = \frac{1+\beta}{2}$，则需 n 满足 $\delta \cdot \sqrt{\frac{n}{p(1-p)}} \geq Z_\beta$，即需要

$n \geq \left(\frac{Z_\beta}{\delta}\right)^2 p(1-p)$。又由：$p(1-p) \leq \frac{1}{4}$，故当 $n \geq \left(\frac{Z_\beta}{2\delta}\right)^2$ 时，取 $n_{\min} = \left[\left(\frac{Z_\beta}{2\delta}\right)^2\right] + 1$，当

$n \geq n_{\min}$ 时，即使：$P\{\omega : |\frac{\mu_n(A)}{n} - p| < \delta\} \geq \beta$

2　一般情形的中心极限定理　（*）

只叙述以下的 *Linderberg* 定理：

定理 3.2　设$\{Y_k，k\geq 1\}$为独立的 *r.v.s.*，记：$a_k=EY_k$, $b_k^2=DY_k<\infty$, $B_n^2 = \sum_{k=1}^{n} b_k^2$

$F_k(x)=P(Y_k\leq x)$，若对$\forall \tau>0$ 有：

$$\lim_{n\to\infty}\frac{1}{B_n}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2dF_k(x)=0 \tag{L}$$

则，对$\forall x\in R$ 有

$$\lim_{n\to\infty}P\{\omega:\frac{1}{B_n}\sum_{k=1}^{n}(Y_k-a_k)\le x\}=\Phi(x). \tag{M}$$

且上式对 x 是一致收敛的。满足(L)式称为 *Linderberg* 条件。

以上定理的证明略。这里仅对(L)条件的概率意义说明如下：

记：$A_k=\{\omega:|Y_k-a_k|>\tau B_n\}$　$1\le k\le n$，则有

$$P\{\omega:\max_{1\le k\le n}|Y_k-a_k|>\tau B_n\}=P(\bigcup_{k=1}^{n}A_k)\le\sum_{k=1}^{n}P(A_k)=\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}dF_k(x)$$

$$\le\frac{1}{(\tau B_n)^2}\sum_{k=1}^{n}\int_{|x-a_k|>\tau B_n}(x-a_k)^2dF_k(x)$$

可见(L)条件保证了，对$\forall\ \tau>0$，

$$\lim_{n\to\infty}P\{\omega:\max_{1\le k\le n}\frac{|Y_k-a_k|}{B_n}>\tau\}=0 \tag{N}$$

这表明：当 n 充分大时，参与构成总和的每一项要以概率“均匀的小”，因而任一被加项对总和极限分布不会导致显著的影响，这样即可是它们总和的分布趋向于正态分布。于是定理可直观的解释为：当某一随机现象的指标是由大量独立的随机因素影响，若这些因素对总和的指标的影响都是“均匀的小”（在概率意义下），则该指标的分布近似于正态分布。这定理深刻的从理论上证明了许多自然界中引出的随机变量可用正态分布来近似的理论根据。

§4 随机变量序列的几种收敛性*

定义 4.1　几乎处处收敛（以概率 1 收敛）。对于 *r.v.s.*$\{X_n, n\ge1\}$，若：

$$P(\omega:\lim_{n\to\infty}X_n(\omega)=X(\omega))=1.$$

则称$\{X_n, n\ge1\}$以概率 1 收敛于 X，或者几乎处处收敛。(*almost everywhere or almost surely* 简记为: *a.e. or a.s.*)。记为：

$$\lim_{n\to\infty}X_n\overset{a.s.}{=}X,\quad \text{或者 } X_n\xrightarrow{a.e.}X \text{ 或者 } X_n\xrightarrow{a.s.}X$$

例 4.1 设$\{X_n, n\ge1\}$，X, Y 是定义在[0,1]上 *Borel* 概率空间（见第一章例 2.16）$(\Omega, \Phi,$

P)=([0,1], B[0,1], P)上的 $r.v.$，满足：$\forall\omega\in[0,1]$，$Y(\omega)=1$。而 $X(\omega)=1$，若$\omega\in\overline{B}$={[0,1]上无理点}；$X(\omega)=0$，若$\omega\in B$={[0,1]上有理点全体}。而 $X_n(\omega)=1$，若$\omega\in(\frac{1}{2^n},1]$；$X_n(\omega)=0$，若 $\omega\in[0,\frac{1}{2^n}]$。则易知：$P(\omega: X(\omega)\neq Y(\omega))=P(B)=0$。$(\omega:\lim\limits_{n\to\infty}X_n(\omega)=Y(\omega))=\Omega$；$(\omega:\lim\limits_{n\to\infty}X_n(\omega)=X(\omega))=\overline{B}\neq\Omega$，但 $P(\overline{B})=1$，故：$\lim\limits_{n\to\infty}X_n\overset{a.s.}{=}X$。

定义 4.2　以概率收敛。对于 $r.v.s.\{X_n,n\geq1\}$，若：$\forall\varepsilon>0$

$$\lim_{n\to\infty}P(\omega:|X_n(\omega)-X(\omega)|<\varepsilon)=1.$$

则称$\{X_n,n\geq1\}$以概率收敛于 X。记为：

$$\lim_{n\to\infty}X_n\overset{P}{=}X,\ 或者\ X_n\xrightarrow{P}X.\quad(n□\infty)$$

例 4.2 设$\{Y_k, 1\leq k\leq n\}$ $i.i.d.$，且 $Y_1\sim U[0,1]$，令 $X_n=\sum\limits_{k=1}^{n}Y_k\Big/n$，则由§1 大数定律可知：$X_n\xrightarrow{P}\frac{1}{2}\quad(n\to\infty)$。

定义 4.3　$L_p(p>0)$收敛。对于 $r.v.s.\{X_n,n\geq1\}$，若：

$$\lim_{n\to\infty}E|X_n-X|^p=0.$$

则称$\{X_n,n\geq1\}$以 L_p 收敛于 X。记为：$X_n\xrightarrow{L_p}X$。

当 $p=2$ 时，$\lim\limits_{n\to\infty}E|X_n-X|^2=0$，称$\{X_n,n\geq1\}$均方收敛到 X (*mean square*)。记为：

$$\lim_{n\to\infty}X_n\overset{m.s.}{=}X,\ 或者\ X_n\xrightarrow{m.s.}X.$$

例 4.3 设$\{X_k, 1\leq k\leq n\}$相互独立，且满足：$P(X_n=1)=\frac{1}{n},P(X_n=0)=\frac{n-1}{n}$，$n\geq1$，$X(\omega)\equiv0$。则 $E(X_n-0)^2=\frac{1}{n}\to0$，故 $\lim\limits_{n\to\infty}E|X_n-0|^2=0$，即 $X_n\xrightarrow{m.s.}0$。

定义 4.4　依分布收敛(弱收敛)。对于 $r.v.s.\{X_n,n\geq1\}$，记：$F_n(x)=P(X_n\leq x)$，$F(x)=P(X\leq x)$分别是 X_n 和 X 的分布函数。若对 $F(x)$的连续点有：

$$\lim_{n\to\infty}F_n(x)=F(x).$$

则称$\{X_n,n\geq1\}$依分布函数收敛于 X (*weak converge*)。记为：

$$X_n\xrightarrow{d}X,或者\ X_n\xrightarrow{w}X$$

例 4.4 Y_n, Z_n 的记号同本章定理 3.1，令 $Z\sim N(0,1^2)$，则 $Z_n\xrightarrow{d}Z$，即$\forall x\in R$，有：

$$\lim_{n\to\infty} P(Z_n \le x) = \Phi(x)\text{。}$$

注：设$\{F_n(x),n\ge1\}$为分布函数序列，若存在单调不减的函数 $F(x)$，使得在 $F(x)$的连续点上有：$\lim\limits_{n\to\infty} F_n(x) = F(x)$，则称 $F_n(x)$弱收敛到 $F(x)$，记为：$F_n(x) \xrightarrow{w} F(x)$。注意，尽管$\{F_n(x),n\ge1\}$是分布函数列，$F(x)$却未必是分布函数。故"$F_n(x)$弱收敛到分布函数 $F(x)$"和"$F_n(x)$弱收敛到函数 $F(x)$"两句话有重大差别。例如：$F_n(x)=0$, 如 $x<n$；$F_n(x)=1$, 如 $x\ge n$；$F(x)=0, x\in(-\infty,+\infty)$。则：$\lim\limits_{n\to\infty} F_n(x) = F(x)$，但 $F(x)$不是分布函数。

以上四种收敛性是常见的收敛定义，为了讨论四种收敛性的关系，先引进一些记号，再给出 *Borel-Cantalli* 引理。

记：$(A_n, i.o.)=\overline{\lim_{n\to\infty}} A_n = \lim\limits_{n\to\infty}\sup A_n \triangleq \bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k$，即：$(A_n, i.o.)=\{\omega : \omega$至少属于无穷多个 $A_n\}$。又记：$\underline{\lim_{n\to\infty}} A_n = \lim\limits_{n\to\infty}\inf A_n \triangleq \bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty} A_k$。

引理 4.1 （***Borel-Cantalli*** **引理**）

（1）若随机事件序列$\{A_n,n\ge1\}$满足$\sum_{n=1}^{\infty} P(A_n) < \infty$，则：$P(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k) = 0.$

（2）若$\{A_n,n\ge1\}$是相互独立的随机事件序列。则：$\sum_{n=1}^{\infty} P(A_n) = \infty$成立的充要条件为

$P(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k) = 1.$

证明：（1）$0\le P(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty} A_k) = P(\lim\limits_{n\to\infty}\bigcup_{k=n}^{\infty} A_k) = \lim\limits_{n\to\infty} P(\bigcup_{k=n}^{\infty} A_k) \le \lim\limits_{n\to\infty}\sum_{k=n}^{\infty} P(A_k) = 0.$

其中第二个等号是利用了概率的连续性，最后一个不等号是因为：$\sum_{n=1}^{\infty} P(A_n) < \infty$。

（2）先证明必要性。由$\{A_n,n\ge1\}$独立，有：$P(\bigcap_{k=n}^{m+n} A_k) = \prod_{k=n}^{m+n} P(A_k)$，故：

$P(\bigcap_{k=n}^{\infty} A_k) = \prod_{k=n}^{\infty} P(A_k)$。又当 $0\le x<1$ 时，$\ln(1-x)\le -x$；故：

$$\ln\prod_{k=n}^{\infty}(1-P(A_k))=\sum_{k=n}^{\infty}\ln(1-P(A_k))\le-\sum_{k=n}^{\infty}P(A_k)=-\infty$$

由此得

$$P(\bigcap_{k=n}^{\infty}\overline{A_k})=0\quad\forall n\ge1\quad\therefore P(\bigcup_{k=n}^{\infty}A_k)=1\quad\forall n\ge1\quad\therefore P(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k)=1.$$

再证明充分性。 若: $P(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k)=1$. 假设：$\sum_{n=1}^{\infty}P(A_n)<\infty$，则由（1）知：

$P(\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k)=0$. 矛盾！ 又 $P(A_n)\ge0$，故只能：$\sum_{n=1}^{\infty}P(A_n)=\infty$ 。

定义 4.5　称 $r.v.s.\{X_n, n\ge1\}$以概率 1 ，或以概率，或以 L_p 是基本列(*Cauchy* 列)若：$P(\omega:\underset{\substack{n\to\infty\\ m\to\infty}}{Lim}(X_n-X_m)=0)=1$，或 $\forall\varepsilon>0, \underset{\substack{n\to\infty\\ m\to\infty}}{Lim}P(\omega:|X_n-X_m|<\varepsilon)=1$，若

$\underset{\substack{n\to\infty\\ m\to\infty}}{Lim}E|X_n-X_m|^p=0.$

以下给出一个序列几乎处处收敛的充要条件。

定理 4.1　（1）$r.v.s.\ \{X_n, n\ge1\}$几乎处处收敛到 X 的充要条件是：$\forall\varepsilon>0$,

$$\underset{n\to\infty}{Lim}P(\omega:\sup_{k\ge n}|X_k-X|\ge\varepsilon)=0$$

（2）$r.v.s.\ \{X_n, n\ge1\}$以概率 1 为基本列的充要条件是：$\forall\varepsilon>0$,

$$\underset{n\to\infty}{Lim}P(\omega:\sup_{\substack{k\ge n\\ l\ge n}}|X_k-X_l|\ge\varepsilon)=0$$

或

$$\underset{n\to\infty}{Lim}P(\omega:\sup_{k\ge1}|X_{n+k}-X_n|\ge\varepsilon)=0$$

证明：（1）记：$A_k(\frac{1}{m})=\left\{\omega:|X_k-X|\ge\frac{1}{m}\right\}$,　$A(\frac{1}{m})=\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k$,

$B=\{\omega:\underset{n\to\infty}{Lim}X_n(\omega)=X(\omega)\}$,　则 $\overline{B}=\{\omega:\underset{n\to\infty}{Lim}X_n(\omega)\ne X(\omega)\}$.

我们有：$\forall\omega\in B,\Leftrightarrow\forall m\ge1,\exists n\ge1,\forall k\ge n,$有：$\omega\in\left\{\omega:|X_k-X|<\frac{1}{m}\right\}\Leftrightarrow\forall m\ge1,\exists n\ge1,$

$\omega\in\bigcap_{k=n}^{\infty}\left\{\omega:|X_k-X|<\frac{1}{m}\right\}\Leftrightarrow\forall m\ge1,\omega\in\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left\{\omega:|X_k-X|<\frac{1}{m}\right\}\Leftrightarrow$

$\omega\in\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left\{\omega:|X_k-X|<\frac{1}{m}\right\}$，故：$B=\bigcap_{m=1}^{\infty}\bigcup_{n=1}^{\infty}\bigcap_{k=n}^{\infty}\left\{\omega:|X_k-X|<\frac{1}{m}\right\}$

$$\therefore \overline{B}=\bigcup_{m=1}^{\infty}A(\frac{1}{m})=\bigcup_{m=1}^{\infty}\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k(\frac{1}{m}).$$

故由 $P(B)=1\Leftrightarrow P(\overline{B})=P(\bigcup_{m=1}^{\infty}A(\frac{1}{m}))=0 \quad\Leftrightarrow\quad P(A(\frac{1}{m}))=0\ \ \forall m\geq 1.$

$\Leftrightarrow\quad P(A(\varepsilon))=0\ \ \forall\varepsilon>0\quad\Leftrightarrow\quad \underset{n\to\infty}{Lim}\,P(\bigcup_{k=n}^{\infty}A_k(\varepsilon))=0$ (概率的连续性)

（注意：$(\omega:\sup_{k\geq n}|X_k-X|\geq\varepsilon)=\bigcup_{k\geq n}A_k(\varepsilon)$）$\Leftrightarrow\ \underset{n\to\infty}{Lim}\,P(\omega:\sup_{k\geq n}|X_k-X|\geq\varepsilon)=0$。

（2）令：$D_{k,l}(\varepsilon)=\{\omega:\ |X_{k+l}-X_l|\geq\varepsilon\}$ $\quad D(\varepsilon)=\bigcap_{n=1}^{\infty}\bigcup_{\substack{k=1\\ l=n}}^{\infty}D_{k,l}(\varepsilon)$

由下列事件等价：$D=\{\omega:\underset{\substack{k\to\infty\\ l\to\infty}}{Lim}(X_k-X_l)\neq 0\}=\bigcup_{\varepsilon>0}D(\varepsilon)=\bigcup_{\varepsilon>0}\bigcap_{n=1}^{\infty}\{\omega:\sup_{\substack{k\geq n\\ l\geq n}}|X_{k+l}-X_l|\geq\varepsilon\}$ 得：

$$P(D)=0\quad\Leftrightarrow\quad P(\bigcup_{\varepsilon>0}D(\varepsilon))=0\quad\Leftrightarrow\quad P(D(\varepsilon))=0\ \ \forall\varepsilon>0.\quad\Leftrightarrow$$

$$P(\bigcap_{n=1}^{\infty}\bigcup_{\substack{k=n\\ l=n}}^{\infty}\{\omega:|X_k-X_l|\geq\varepsilon\})=\underset{n\to\infty}{Lim}\,P(\bigcup_{\substack{k=1\\ l=n}}^{\infty}\{\omega:|X_{k+l}-X_l|\geq\varepsilon\})=0\quad\Leftrightarrow$$

$$\underset{n\to\infty}{Lim}\,P(\omega:\sup_{\substack{k\geq n\\ l\geq n}}|X_k-X_l|\geq\varepsilon)=0\quad\forall\varepsilon>0$$

推论 1：$X_n\xrightarrow{a.e.}X\ (n\to\infty)$ 的充分条件是：$\forall\varepsilon>0,\ \sum_{k=1}^{\infty}P\{\omega:|X_k-X|\geq\varepsilon\}<\infty$

证明：因 $P\{\omega:\sup_{k\geq n}|X_k-X|\geq\varepsilon\}=P\{\bigcup_{k=n}^{\infty}(\omega:|X_k-X|\geq\varepsilon)\}<\sum_{k=1}^{\infty}P\{\omega:|X_k-X|\geq\varepsilon\}$

再由定理即知推论成立。

推论 2：设 $\varepsilon_n>0,\varepsilon_n\to 0\quad(n\to\infty)$，若 $\sum_{n=1}^{\infty}P\{\omega:|X_n-X|\geq\varepsilon_n)<\infty$，则：

$$X_n\xrightarrow{a.e.}X.$$

证明：令 $A_n=\{\omega:|X_n-X|\geq\varepsilon_n\}$ 则 $P\{\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k\}=0$，这意味着几乎对所有的 $\omega\in$ Ω，存在 $N=N(\omega)$，对 $\forall n\geq N(\omega)$ $|X_n-X|\leq\varepsilon_n$，但 $\varepsilon_n\to 0$，故：

$$P\{\omega:\underset{n\to\infty}{Lim}X_n\neq X\}=P\{\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k\}=0\ .$$

得：$X_n\xrightarrow{a.e.}X$

以下定理将给出几种收敛性的关系

定理 4.2

（1）$X_n\xrightarrow{a.e.}X\Rightarrow X_n\xrightarrow{P}X$，

（2）$X_n\xrightarrow{L_p}X\Rightarrow X_n\xrightarrow{P}X$，

（3）$X_n\xrightarrow{L_p}X\Rightarrow X_n\xrightarrow{d}X$ 。

证明：（1）由定理 5.1 的（1）及定义而得。（2）由 *Chebyshev* 不等式知：$\forall\varepsilon\geq 0$

$P\{\omega:|X_n-X|\geq\varepsilon\}\leq\dfrac{E|X_n-X|^p}{\varepsilon^p}$，由 $X_n\xrightarrow{L_p}X$，即：$E|X_n-X|^p\to 0\quad(n\to\infty)$

故 $\forall\varepsilon\geq 0,P\{\omega:|X_n-X|\geq\varepsilon\}\to 0\quad(n\to\infty)$ 即 $X_n\xrightarrow{P}X$ 。

（3）记 $F(x)=P(X\leq x)$，$F_n(x)=P(X_n\leq x)$，$\forall x,y\in\mathrm{R}$，

$(X\leq y)=(X_n\leq x,X\leq y)\bigcup(X_n>x,X\leq y)\subset(X_n\leq x)\bigcup(X_n>x,X\leq y)$，$\Rightarrow F(y)\leq$ $F_n(x)+P(|X_n-X|\geq x-y)$。当 $X_n\xrightarrow{P}X$，可得：$\forall y<x\quad P\{\omega:X_n>x,X\leq y\}\leq$

$P\{\omega:|X_n-X|\geq x-y\}\to 0\quad(n\to\infty)$，　$\therefore F(y)\leq\underset{n\to\infty}{\underline{Lim}}F_n(x)$。

类似的有：

$\forall x<z\quad(X_n\leq x)=(X_n\leq x,X\leq z)\bigcup(X_n\leq x,X>z)\subset(X\leq z)\bigcup(X_n\leq x,X>z)$，得

$Fn(x)\leq F(z)+P(|Xn-X|\geq z-x)$。

又 $P\{\omega:X_n\leq x,X>z\}\leq P\{\omega:|X-X_n|\geq z-x\}\to 0\quad(n\to\infty)$，$\therefore\underset{n\to\infty}{\overline{Lim}}F_n(x)\leq F(z)$

对 $y<x<z$ 有：$F(y) \le \underset{n\to\infty}{\underline{Lim}} F_n(x) \le \underset{n\to\infty}{\overline{Lim}} F_n(x) \le F(z)$

如 x 是 $F(x)$的连续点，令 $y\uparrow x, z\downarrow x$ 得：$F(x) = \underset{n\to\infty}{Lim} F_n(x).$

定理 4.2 的（1）、(2)、(3）其逆不真, 以下举例说明之：

例 4.5 $X_n \xrightarrow{P} X \quad \not\Rightarrow \quad X_n \xrightarrow{a.e.} X$ 。

令 (Ω, Φ, P)=([0,1], B[0,1], P)为[0,1]上 *Borel* 概率测度（见第一章例 2.16），令：

$$\begin{aligned}
&X_1 = I_{(0,1]},\\
&X_2 = I_{(0,\,1/2]}, X_3 = I_{(1/2,1]},\\
&X_4 = I_{(0,1/3]}, X_5 = I_{(1/3,\,2/3]}, X_6 = I_{(2/3,1]},\\
&X_7 = I_{(0,1/4]}, \qquad \cdots\cdots \qquad X_{10} = I_{(3/4,1]},\\
&\cdots\cdots
\end{aligned}$$

对任意正整数 $n\ge1$，总存在二正整数 i，k，$1\le i\le k$，$k\ge1$ 使 $n=(k-1)k/2+i$。故对$\forall$ $n\ge1$，有 $X_n = X_{(k-1)k/2+i} = I_{((i-1/k,i/k)]}, 1\le i\le k, k\ge1$， 则有 $EX_n = 1/k, EX_n^2 = 1/k$, $n=(k-1)k/2+i$

$1\le i\le k$，$k\ge1$。$\forall\,\varepsilon>0$，$P(|X_n(\omega)|\ge\varepsilon)\to 0, (n\to\infty)$，故 $X_n \xrightarrow{P} 0$ 。

但是对$\forall\ \omega\in[0,1]$，$\forall k\ge1$，因 $\sum_{i=1}^{k} I_{((i-1)/k,i/k]} = I_{(0,1]}$，故对$\forall\ \omega\in[0,1]$及$\forall k\ge1$ 存在 $i_k(1\le i_k\le k)$，有 $I_{((i_k-1)/k,i_k/k]} = 1$，即存在 $n=(k-1)k/2+i_k$，使 $X_n(\omega)=1$，而对其余的 $i\ne i_k(1\le i\le k)$，有 $X_n(\omega)=0$。由此可知，对$\forall\ \omega\in[0,1]$，序列$\{X_n(\omega), n\ge1\}$中有无穷多个 $X_n(\omega)=1$ 及有无穷多个 $X_n(\omega)=0$ 。于是 $X_n(\omega)$对每一个 $\omega\in[0,1]$都不收敛。

例 4.6 $X_n \xrightarrow{a.e.} X \quad \Rightarrow X_n \xrightarrow{P} X \quad \not\Rightarrow X_n \xrightarrow{L_p} X$

设 (Ω, Φ, P)=([0,1], B[0,1], P)为[0,1]上 *Borel* 概率测度（见第一章例 2.16），令：

$$X_n(\omega) = \begin{cases} e^n & 0\le\omega\le\dfrac{1}{n} \\ 0 & \dfrac{1}{n}<\omega\le1. \end{cases}$$

易见 $P(\omega: \underset{n\to\infty}{Lim} X_n \ne 0) = 0.\quad \forall\varepsilon>0, P(\omega:|X_n(\omega)|<\varepsilon) = 1-\dfrac{1}{n}\to1 \quad (n\to\infty)$

则有：$X_n \xrightarrow{a.e.} 0.\quad X_n \xrightarrow{P} 0.$ 但 $E|X_n|^p = \dfrac{e^{np}}{n}\to\infty$ 即 X_n 不以 L_p 收敛到 0。

例 4.7 $X_n \xrightarrow{L_p} X \not\Rightarrow \quad X_n \xrightarrow{a.e.} X$

设 $r.v.s.\{X_n, n\geq 1\}$，满足：$P(X_n=1)=\dfrac{1}{n}, P(X_n=0)=\dfrac{n-1}{n}$，且$\{X_n, n\geq 1\}$独立。易证明 $X_n \xrightarrow{L_p} 0 \quad (n\to\infty)$，对每一个 $p>0$。但 $\sum\limits_{n=1}^{\infty} P(|X_n-0|>\dfrac{1}{2})=\sum\limits_{n=1}^{\infty}\dfrac{1}{n}=\infty$，故由 *Borel—Cantalli* 引理 X_n 不以概率 1 收敛到 0，即 $\mathop{Lim}\limits_{n\to\infty} X_n \overset{a.e.}{\neq} 0$。

尽管一般由 $X_n \xrightarrow{P} X$ 不能推出 $X_n \xrightarrow{a.e.} X$，但有如下的：

定理 4.3 设 $r.v.s.\{X_n, n\geq 1\}$和 $r.v. X$ 有 $X_n \xrightarrow{P} X$ 则存在一子序列 $\{X_{n_k}, k\geq 1\}$ 使

$$X_{n_k} \xrightarrow{a.e.} X \quad (k\to\infty)。$$

证明：由 $X_n \xrightarrow{P} X$ 的定义可知，若令 $n_1=1$，则对每一个 $k\in \boldsymbol{N}, \exists n_k$，且 $n_k>n_{k-1}$ 使

$P\{\omega:|X_{n_k}-X|\geq \dfrac{1}{2^k}\}<\dfrac{1}{2^k}\ \therefore \forall\, 0<\varepsilon<1$, 及一切$k\geq N_0=-\dfrac{\ln\varepsilon}{\ln 2} \quad (\dfrac{1}{2^k}\leq\varepsilon)$，与 $N\geq N_0$

有 $\sum\limits_{k=N}^{\infty} P\{\omega:|X_{n_k}-X|\geq\varepsilon\}\leq \sum\limits_{k=N}^{\infty} P\{\omega:|X_{n_k}-X|\geq \dfrac{1}{2^k}\}<\sum\limits_{k=N}^{\infty}\dfrac{1}{2^k}=\dfrac{1}{2^{N+1}}$

$\therefore \sum\limits_{k=1}^{\infty} P\{\omega:|X_{n_k}-X|\geq\varepsilon\}=\sum\limits_{k=1}^{N_0} P\{\omega:|X_{n_k}-X|\geq\varepsilon\}+\sum\limits_{k=N_0}^{\infty} P\{\omega:|X_{n_k}-X|\geq\varepsilon\}$

$\leq N_0+\dfrac{1}{2^{N_0+1}}<\infty$。最后一个不等号是因为：对于取定的ε>0，$N_0=-\dfrac{\ln\varepsilon}{\ln 2}<\infty$。

故 ∀ε>0，$\sum\limits_{k=1}^{\infty} P\{\omega:|X_{n_k}-X|\geq\varepsilon\}<\infty$。由推论 1 知定理成立。

§5 强大数定律*

本节讨论关于独立随机变量序列的前 n 项算术平均以概率 1 收敛到它们的数学期望（理论平均）的平均的一些基本结果。通常称为强大数定律。

1　*Borel* 强大数定律

定理 5.1 设在 *Bernoulli* 独立重复试验序列中，事件 A 出现的概率为 $P(A)=p$，以$\mu_n(A)$表示前 n 次试验中 A 出现的次数。则：

$$P(\omega:\mathop{Lim}\limits_{n\to\infty}\frac{\mu(A)}{n}=p)=1 \tag{5.1}$$

证明： (1)成立的充要条件是：$\forall\varepsilon>0$

$$P\{\omega:\bigcap_{l=1}^{\infty}\bigcup_{n=l}^{\infty}|\frac{\mu_n(A)}{n}-p|\geq\varepsilon\}=0 \tag{5.2}$$

若记 $A_n(\varepsilon)=\{\omega:|\frac{\mu_n(A)}{n}-p|\geq\varepsilon\}$，(3.2)式为：$\forall\varepsilon>0$

$$P\{\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k(\varepsilon)\}=0. \tag{5.3}$$

故由 *Borel-Cantalli* 引理，只需证 $\sum_{n=1}^{\infty}P\{A_n(\varepsilon)\}<\infty$。为此记：$X_n=I_{A_k}$，$A_k$表示第 k 次试验出 A 的事件，则：$\frac{\mu_n(A)}{n}-p=\frac{1}{n}\sum_{k=1}^{n}(X_k-p)$。由 *Chebyshev* 不等式的推广(*Markov* 不等式)。有：$P\{A_n(\varepsilon)\}\leq\frac{1}{\varepsilon^4}E|\frac{\mu_n(A)}{n}-p|^4$

又 $E|\frac{\mu_n(A)}{n}-p|^4=\frac{1}{n^4}\cdot E[\sum_{k=1}^{n}(X_k-p)]^4$

$=\frac{1}{n^4}\cdot E\{[\sum_{k=1}^{n}(X_k-p)]\cdot[\sum_{l=1}^{n}(X_l-p)]\cdot[\sum_{i=1}^{n}(X_i-p)]\cdot[\sum_{j=1}^{n}(X_j-p)]\}$

$=\frac{1}{n^4}\cdot\{\sum_{i=1}^{n}E(X_i-p)^4+\sum\sum_{i\neq k}E\{(X_i-p)^2\cdot(X_k-p)^2\}$

在和式中注意到 $E(X_i-p)=0$，因而上式中只有 $E(X_i-p)^4$ 及 $E\{(X_i-p)^2\cdot(X_k-p)^2\}$ $(k\neq i)$ 不为 0。而 $E(X_i-p)^4$ 有 n 项，$E\{(X_i-p)^2\cdot(X_k-p)^2\}$。有 $C_4^2\cdot C_n^2=3\cdot n(n-1)$ 项，又：$E(X_i-p)^4=pq(p^3+q^3)$，且 $E\{(X_i-p)^2\cdot(X_k-p)^2\}=p^2\cdot q^2,(i\neq k)$，其中 $q=1-p$。

故

$E|\frac{\mu_n(A)}{n}-p|^4=\frac{pq}{n^4}\cdot[npq(p^3+q^3)+3pqn(n-1)]\leq\frac{pq}{n^3}[p^3+q^3+3pq(n-1)]$

$<\frac{pq}{n^3}[n(p^3+q^3)+n\cdot 3pq]\leq\frac{1}{n^2}\cdot\frac{1}{4}\cdot[p^3+q^3+3pq(p+q)]=\frac{1}{4n^2}$

最后一个不等号是因为 $pq\leq\frac{1}{4}$，故

$$P(A_n(\varepsilon))\leq\frac{1}{4\varepsilon^4}\cdot\frac{1}{n^2}\quad\Rightarrow\quad\sum_{n=1}^{\infty}P(A_n(\varepsilon))\leq\frac{1}{4\varepsilon^4}\sum_{n=1}^{\infty}\frac{1}{n^2}<\infty\Rightarrow P\{\bigcap_{n=1}^{\infty}\bigcup_{k=n}^{\infty}A_k(\varepsilon)\}=0.$$

$$\therefore P\{\omega:\lim_{n\to\infty}\frac{\mu(A)}{n}=p\}=1。$$

例 5.1 （强大数定律应用之一）经验分布以概率 1 收敛到理论分布

X~$F(x)$，$\{X_k, 1\le k\le n\}$ $i.i.d.$及 $\widetilde{F}_n(x)$ 同例 1.1，应用定理 5.1，对$\forall x\in R$，有：

$$P\{\underset{n\to\infty}{Lim}\widetilde{F}_n(x)=F(x))=1.$$

注: ***Markov* 不等式** 对 $r.v.X$，若 $E|X-EX|^r$ 存在(r>0 常数)，则$\forall\varepsilon>0$

$$P\{\omega:|X-EX|\ge\varepsilon\}\le\frac{E|X-EX|^r}{\varepsilon^r}\quad(r>0)$$

可仿照 *Chebyshev* 不等式的证明加以证明，留给读者。

2　*Cantalli* 强大数定律

定理 5.2 设$\{Y_n,n\ge1\}$为独立的 $r.v.s.$,且有有限的四阶矩，即存在常数 C

$E|Y_n-EY_n|^4<C\quad\forall n\ge1$。令：$X_n=\frac{1}{n}\sum_{k=1}^{n}Y_k$，则：

$$P\{\omega:\underset{n\to\infty}{Lim}(X_n-EX_n)=0\}=1\qquad(5.4)$$

证明：不是一般性令 EY_n=0；由 *Borel-Cantelli* 引理知，为证(4.3)式成立只需证明：

$\forall\varepsilon>0$，$\sum_{n=1}^{\infty}P\{\omega:\frac{1}{n}\cdot|\sum_{k=1}^{n}Y_k-\sum_{k=1}^{n}EY_k|\ge\varepsilon\}<\infty.$

为应用 *Chebyshev* 不等式，只需证明：$\sum_{n=1}^{\infty}E|\frac{1}{n}\sum_{k=1}^{n}Y_k|^4<\infty$。

注意到：$EY_i=0$，所以$E(\sum_{k=1}^{n}Y_k)^4=\sum_{i=1}^{n}EY_i^4+6\cdot\sum_{\substack{i,j=1\\ i\ne j}}^{n}EY_i^2EY_j^2\le nC+6\sum_{\substack{i,j=1\\ j\ne i}}^{n}\sqrt{EY_i^4\cdot EY_j^4}$

$\le nC+\frac{6n(n-1)}{2}C<3n^2C$

所以 $\frac{1}{n^4}\cdot E(\sum_{k=1}^{n}Y_k)^4\le\frac{3C}{n^2}\Rightarrow\sum_{n=1}^{\infty}E|\frac{1}{\mathrm{n}}\sum_{k=1}^{n}Y_k|^4\le3C\cdot\sum_{n=1}^{\infty}\frac{1}{n^2}<\infty$。

3 .*Kolmogorov* 强大数定律

（1）*Kolmogorov* 不等式

定理 5.3 设$\{Y_n,n\geq 1\}$为独立的 *r.v.s.*，$DY_n=\sigma_n<\infty$，令 $X_n=\frac{1}{n}\sum_{k=1}^{n}Y_k$ ，则对 $\forall\varepsilon>0$，

$$P\{\omega:\max_{1\leq k\leq n}|X_k-EX_k|\geq\varepsilon\}\leq\frac{DX_n}{\varepsilon^2} \tag{5.5}$$

这是 *Chebyshev* 不等式的推广，是更精细的不等式，它是研究收敛性的重要工具。

（2）*Kolmogorov* 强大数定律

定理 5.4 设$\{Y_n,n\geq 1\}$为独立的 *r.v.s.*，且$\sum_{n=1}^{\infty}\frac{DY_n}{n^2}<\infty$，则

$$P\{\omega:\mathop{Lim}_{n\to\infty}\frac{1}{n}\cdot\sum_{k=1}^{n}(Y_k-EY_k)=0\}=1. \tag{5.6}$$

以上结果的证明已超出本书的范围，故从略。

练习题

5.1 设$X_1,\cdots,X_{50}$独立同参数$(\lambda=0.02)$的泊松分布，记$Y=\sum_1^{50}X_i$，试利用中心极限定理近似计算$P(Y\geq 3)$。

5.2 计算机在进行加法时，对每个加数取整，设所有的取整误差是相互独立且同分布于$U(-0.5,0.5)$；(1)若将1500个数相加，问误差总和的绝对值小于15的概率是多少？（2）几个数加在一起可使得误差总和的绝对值小于10的概率为0.90。

5.3 设有30个同类型的电子器件$D_1,D_2,\cdots,D_{30}$,它们的使用情况如下：D_1损坏，D_2立即使用，D_2损坏，D_3立即使用，等等。设它们的寿命是独立同参数（a=1小时）的指数分布，令T为30个器件的总寿命，问：T超过350小时的概率是多少？

5.4 设$r.v.s.\{X_n,n\geq 1\},X_n\xrightarrow{P}X,X_n\xrightarrow{P}Y$。证：$P\{\omega:X=Y\}=1$。

5.5 设$r.v.s.\{X_n,n\geq 1\},X_n\xrightarrow{P}X,Y_n\xrightarrow{P}Y$。证：

$X_n\pm Y_n\xrightarrow{P}X\pm Y,|X_n|\xrightarrow{P}|X|$，$X_n\cdot Y_n\xrightarrow{P}X\cdot Y$。

5.6 若$X_n\xrightarrow{d}C$（常数），证：$X_n\xrightarrow{P}C$。

5.7 设$\sum_{n=1}^{\infty}E|X_n|^p<+\infty$, 对$p>0$，证：$X_n\xrightarrow{a.e.}0$。

5.8 设$\{F_n(x)\}$为正态分布函数列，收敛于分布函数$F(x)$。证：$F(x)$也是正态分布函数。

5.9 设分布函数列$\{F_n(x)\}$弱收敛到连续的分布函数$F(x)$。试证：这收敛对于$x\in R$是一致的。

5.10 设 $r.v.$序列$\{X_n,n\geq 1\}$及$X,\forall m,k\in N$，记$A_k(m)=\left\{|X_k-X|\geq\frac{1}{m}\right\}$，若对$\forall m\in N$有$\sum_{k=1}^{\infty}P(A_k(m))<+\infty$，证：$P(\lim\limits_{n\to\infty}X_n=X)=1$。

5.11 设$\{X_i,1\leq i\leq n\}$ $i.i.d.$，$X_n\sim U$（0,1），对$\forall 0<t<1$，记$N_n(t)=\sum_{i=1}^{n}I_{(X_i\leq t)}$，

$F_n(t)=\frac{1}{n}N_n(t),\alpha_n(t)=\sqrt{n}(F_n(t)-t)$。证：(1) $\forall\varepsilon>0,\ \underset{n\to\infty}{Lim}P(|F_n(t)-t|<\varepsilon)=1$；

(2) $\forall x\in R,0<s<1,\ \lim\limits_{n\to\infty}P(\alpha_n(s)\le x(s(1-s))^{1/2})=\Phi(x)$。

5.12 设$\{X_k,1\le k\le n\}$独立，且$X_k\sim Ex(\lambda)$，$\lambda=\frac{1}{\theta}>0$即$X_k$的分布函数为$F(t)=(1-e^{-\lambda t})I_{(t\ge 0)}$，$X_{(1)},X_{(2)},\cdots,X_{(n)}$为其顺序统计量，$\forall t\ge 0,N_n(t)=\sum_{k=1}^n I_{(X_{(k)}\le t)}$，

$\hat{F}_n(t)=N_n(t)/n,\ \hat{\theta}_{n,k}=\left(\sum_{i=1}^k X_{(i)}+(n-k)X_{(k)}\right)/k,1\le k\le n$，证：

（1）$\forall\varepsilon>0,$有$\lim\limits_{n\to\infty}P\left(\left|\hat{F}_n(t)-F(x)\right|<\varepsilon\right)=1$；

（2）$\forall\varepsilon>0,$有$\lim\limits_{n\to\infty}P\left(\left|\frac{1}{n}\sum_{i=1}^n X_{(i)}-\theta\right|<\varepsilon\right)=1$；

（3）$\forall x\in R,\ \lim\limits_{n\to\infty}P\left(\sqrt{n}(\hat{\theta}_{n,n}-\theta)\Big/\sqrt{\theta}\le x\right)=\Phi(x)$，其中$\Phi(x)$为标准正态分布函数。（提示：参见 4.28）

5.13 设$r.v.X\ge 0,$且$P(X<\infty)=1$，令$X_n=\sum_{k=0}^{n2^n-1}\frac{k}{2^n}I_{(\frac{k}{2^n}\le X<\frac{k+1}{2^n})}+nI_{(X\ge n)}$

证：$\lim\limits_{n\to\infty}X_n=X(a.s.)$。

5.14 设$r.v.X\ge 0,$且$P(X<\infty)=1$，记$X^+=X\vee 0,X^-=-(X\wedge 0)$，令

$X_n^+=\sum_{k=0}^{n2^n-1}\frac{k}{2^n}I_{(\frac{k}{2^n}\le X^+<\frac{k+1}{2^n})}+nI_{(X^+\ge n)}$，$X_n^-=\sum_{k=0}^{n2^n-1}\frac{k}{2^n}I_{(\frac{k}{2^n}\le X^-<\frac{k+1}{2^n})}+nI_{(X^-\ge n)}$

$X_n=X_n^+-X_n^-$。证：$\lim\limits_{n\to\infty}X_n=X(a.s.)$。

第六章 泊松信号流

§0 随机过程的有关概念

在第二章~第四章中，我们研究了一个随机变量和 n 维随机向量有关的基本概念、理论与方法。第五章的极限理论，涉及了相互独立随机变量序列的极限性态。但是要描述一个随机动态系统问题的概率性质，必须将上述情形加以推广，进而研究无穷多个相互有关的随机变量族。并将它们作为整体加以刻划与研究。

1 随机过程的概念

设对一个参数 $t\in T$，$X(t,\omega)$是一随机变量，称随机变量族 $X_T \hat{=} \{X(t,\omega),t\in T\}$ 为一随机过程（*Stochastic Processes*）。其中 $T\subset \mathrm{R}$，称为指标集。用函数（映射）的观点来描述 X_T，$X(t,\omega):T\times\Omega\to \mathrm{R}$，即 $X(\cdot,\cdot)$ 是定义在 $T\times\Omega$ 上，取值在 R 上的二元单值函数。固定 $t\in T$，$X(t,\cdot)$ 是随机变量；对固定的 $\omega\in\Omega$，$X(t,\omega)$ 是参数 t 的一般函数，称它是随机过程的一条（样本）轨道，或称为一个实现。记号 $X(t,\omega)$ 有时记为 $X_t(\omega)$，简记为：$X(t)$ 或 X_t。

通常参数集 $T\subset R$ 有 3 种：（i）$T=N_0=\{0,1,2,\cdots\}$；（ii）$T=Z=\{0,\pm1,\pm2,\cdots\}$；（$iii$）$T=[a,b]\subset R$ 或 $T=R$。

当 T 为可列集时，称 X_T 为随机序列。X_T 的取值全体之集，称为状态空间，记作 S。S 可以是 N_0、Z、R、R^n 或 C^n 或一般抽象空间。

2 几个例子

（1）质点在直线整数点无限制的随机游动（简单随机游动）。

设一质点在开始 0 时刻在 i（整数）位置，以后每隔单位时间分别以 p 及 q=1−p 向正（右）或负（左）方向随机游动一单位，且各次游动相互独立，如此无穷多次重复游动下去，称为一随机试验。记 X_n 为质点在 n 时刻的位置，固定 n，X_n 是 $r.v.$。对不同的 $n\ge 0$，

$\{X_n, n \geq 0\}$ 是一随机序列。若记 $S_n(1)$为 n 时刻之后，质点到达 1 位置的次数；对不同的 n，$\{S_n(1), n \geq 1\}$ 是一随机过程。

如果我们要考察 $S_1(1)$，$S_2(1)$，…的概率特性，还要考察 $\{S_n(1), n \geq 1\}$ 与 $\{X_n, n \geq 0\}$ 的关系，这时只限在单个的 X_n 或单个的 $S_n(1)$考虑是不够的，需要将 $\{X_n, n \geq 0\}$ 与 $\{S_n(1), n \geq 1\}$ 作为整体来研究。

作为练习，对给定的一次试验，试画出 $n - X_n$ 的图形。

（**2**）考虑某"服务站"在[0，t]上到达的"顾客"数，记为 $N(t)$。固定 t，$N(t)$是一 *r.v.*。给定 $0<s<t$，$N(s), N(t)$是二个 *r.v.*，因此 $\{N(t), t \geq 0\}$ 是一随机过程。又记 S_n 为第 n 个顾客到达时刻，自然我们要关心 S_1，S_2，…，S_n 的情况以及它们与 $N(t)$的关系。这时需要对二个过程 $\{N(t), t \geq 0\}$ 与 $\{S_n, n \geq 1\}$ 作为整体来研究其概率特性。

（**3**）考虑最简单的 R—C 电路的输入输出系统，该系统有一个干扰信号电压，记为 $\xi(t)$；记 $Q(t)$为 t 时电路的电容，它满足：$R\dfrac{dQ(t)}{dt} + \dfrac{1}{C}Q(t) = \xi(t)$。若 $\{\xi(t), t \geq 0\}$ 是一随机过程，则 $\{Q(t), t \geq 0\}$ 也是一个随机过程。

3 有限维分布族与柯尔莫哥洛夫（*Kolmogorov*）定理

对于随机过程 $X_T = \{X(t), t \in T\}$ 如何刻划它的概率特性呢？显然，我们要对每个 $t \in T$ 知道 X_T 的分布（函数）。$F(t;x) \triangleq P(X(t) \leq x)$，称 $F(t;x)$ 为 X_T 的**一维分布**。对于 $\forall t_1, t_2 \in T$，要刻划 $(X(t_1), X(t_2))$ 的特性，就需要用它们的二维分布：$F(t_1, t_2; x_1, x_2)$ $\triangleq P(X_1(t_1) \leq x_1, X_2(t_2) \leq x_2)$。一般的，对 $\forall t_k \in T, 1 \leq k \leq n$，称

$$F(t_1, \cdots, t_n; x_1, \cdots, x_n) \triangleq P(X(t_1) \leq x_1, \cdots, X(t_n) \leq x_n)$$

为 X_T 的 n 维分布。用以描述 n 时刻的概率。其全体

$$\{F(t_1, \cdots, t_n; x_1, \cdots, x_n), t_k \in T, 1 \leq k \leq n, n \geq 1\}$$

称为随机过程 X_T 的有限维分布族。容易看出，它具有以下两个性质：

（1）对称性

对(1,2,…,n)的任一种排列($j_1, j_2, \cdots, j_n$)，均有

$$F(t_1, \cdots, t_n; x_1, \cdots, x_n) = F(t_{j_1}, \cdots, t_{j_n}; x_{j_1}, \cdots, x_{j_n})$$

（2）相容性

对 $m<n$ 有

$$F(t_1,\cdots,t_m,\cdots,t_n;x_1,\cdots,x_m,\infty,\cdots,\infty)=F(t_1,\cdots,t_m;x_1,\cdots,x_m)$$

自然要问，一个随机过程 X_T 的有限维分布族，是否描述了该过程的全部概率特性？下面的定理给出了肯定的回答。

定理（柯尔莫哥洛夫存在性定理）

设分布函数族 $\{F(t_1,\cdots,t_n;x_1,\cdots,x_n),t_k\in T,1\le k\le n,n\ge 1\}$ 满足以上对称性和相容性，则必有一随机过程 $X_T=\{X(t),t\in T\}$，使

$$\{F(t_1,\cdots,t_n;x_1,\cdots,x_n),t_k\in T,1\le k\le n,n\ge 1\}$$

恰好是 X_T 的有限维分布族，即

$$F(t_1,\cdots,t_n;x_1,\cdots,x_n)=P(X(t_1)\le x_1,\cdots,X(t_n)\le x_n)$$

定理证明从略。（有兴趣可参见王梓坤，《随机过程论》，科学出版社 1965）。这说明 X_T 的有限维分布族包含了 X_T 的所有概率信息。

4 数字特征

对随机过程，自然可以利用前面介绍的数字特征来刻划过程某些时刻的局部性质。

称 $m(t)=EX(t)$，为 X_T 的均值函数；$D(t)\triangleq E(X(t)-m(t))^2$ 为其方差函数；$R(s,t)=\mathrm{cov}(X(s),X(t))$ 为其协方差函数。$\rho(s,t)=\mathrm{cov}(X(s),X(t))\cdot(DX(s)\cdot DX(t))^{-1/2}$ 为其相关函数。

本教材六～九章着重介绍几个最简单、最基本、最具特色、最重要，且在应用中越来越显示其威力，理论研究上越来越有魅力的随机过程。特别是着力于介绍具有特殊性质的随机过程，如何抓住其特点，引出相应的概念与工具，从而使刻划与研究它们变得简明、深入且便于应用是我们讨论的重点。

本章主要讨论 *Poisson* 过程及其推广。

§1 泊松信号流的定义

在历史上泊松分布是作为二项分布的近似，于 1837 年由法国数学家泊松(*Poisson*)引入的，故得此名。近数十年来，与泊松分布紧密相关的泊松过程日益显示其重要性，成为随机过程中最重要的过程之一。其主要原因有下面两点：一是许多随机现象可用泊松过程来描述，诸如在给定一时间区间内，电话交换台中来到的呼叫流，公共汽车站来到的乘客流，计数器纪录的到达信号流；又如放射性分裂落到某区域的质点流，热电子的发射流，宇宙中的行星流等等都要以泊松过程及其衍生过程作为主角；二是对泊松过程的深入研究（特别是通过随机过程的研究）已发现它具有许多特殊的优良性质和作用。它是研究其它复杂随机现象及过程的最重要的基础之一。它在信息网络、电子学、医学、物理、生物、宇宙学、环保及管理科学等领域有广泛的应用。

以 $N(t)$表示$(0,t\,]$上来的信号数，$N(s+t)-N(s)$代表$(s,s+t]$上来的信号数。显然，它取值非负整数，且对$\forall t_1<t_2$，都有 $N(t_1)\le N(t_2)$。称 $\{N(t),t\ge 0\}$ 为计数过程。

定义 1.1 若信号流$\{N(t), t\ge 0\}$满足下面四条性质：

ⅰ) $N(0)=0$,

ⅱ) 增量独立性：$\forall 0\le t_1<t_2<\cdots<t_n$； $N(t_1)-N(0), N(t_2)-N(t_1),\cdots,N(t_n)-N(t_{n-1})$

是相互独立的。即，在互不重叠的时间区间上到达的信号数相互独立。

ⅲ) 增量平稳性：$\forall s,t>0, n\in \mathrm{N}$, $P(N(s+t)-N(s)=n)=P(N(t)=n)$

即，在$(s,s+t]$内到达的信号数的概率只与时间间隔长短有关而与时间起点 s 无关。

ⅳ)普通性：在充分小的时间间隔 h 中，最多到达一个信号，来两个或两个以上信号的概率是 h 的高阶无穷小。即$\forall t\ge 0$ ，充分小的 $h>0$，有

$$\begin{aligned}&P(N(t+h)-N(t)=1)=\lambda\cdot h+o(h)\quad(\lambda>0)\\&P(N(t+h)-N(t)\ge 2)=o(h)\end{aligned}\text{，其中 } o(h)\text{：}\lim_{h\to 0}\frac{o(h)}{h}=0$$

则称$\{N(t),t\ge 0\}$是参数为λ 的简单泊松(*Poisson*)信号流，简称 *Possion* 流。（简记：*p.p.*）

注：对于固定的 $t\ge 0$, $N(t)$ 是一个随机变量。而$\{N(t),t\ge 0\}$ 通常表示与参数 t 有关的一族随机变量。

下面我们来推导 $N(t)$的分布。

定理 1.1 如果$\{N(t),t\ge 0\}$是简单泊松流，那么对于$\forall s,t\ge 0$ ，有

$$P(N(s+t)-N(s)=n)=\frac{(\lambda t)^n}{n!}e^{-\lambda t}\quad(n=0,1,2\cdots)\tag{1.1}$$

即$(N(s+t)-N(s))\sim Po(\lambda t)$。

证明：令 $P_n(t)=P(N(s+t)-N(s)=n)$，由增量平稳性可得 $P_n(t)=P(N(t)=n)$。为证(1.1)式，需考察 $P_n(t)$随 t 变化的具体规律，这就要分析在$[0,t+h]$上到达的信号数与$[0,t]$上到达的信号数的关系。取充分小的 $h>0$，事件可分解为：$n\ge 1$ 时

$$(N(t+h)=n)=(N(t)=n, N(t+h)-N(t)=0)$$
$$\bigcup(N(t)=n-1, N(t+h)-N(t)=1) \tag{1.2}$$
$$\bigcup_{k=2}^{n}(N(t)=n-k, N(t+h)-N(t)=k)$$

当 n=0 时，有

$$(N(t+h)=0)=(N(t)=0, N(t+h)-N(t)=0)$$

由增量独立性可得

$$P_0(t+h)=P_0(t)P_0(h)$$

再由普通性可知

$$P(N(t+h)-N(t)=0)=P_0(h)=1-\lambda h+o(h)$$

故

$$P_0(t+h)=P_0(t)(1-\lambda h+o(h))$$

移项，两边同除以 h，得

$$\frac{P_0(t+h)-P_0(t)}{h}=\frac{-\lambda h P_0(t)+o(h)}{h}$$

令 $h\to 0$，上式两边取极限，及 $N(0)$=0，有

$$\begin{cases} P_0'(t)=-\lambda P_0(t) \\ P_0(0)=P(N(0)=0)=1 \end{cases} \tag{1.3}$$

解上面的微分方程(1.3) ，可得 $P_0(t)=e^{-\lambda t}$ 。因此当 n=0 时，定理 1.1 成立。

下面来证 $n\geq 1$ 时的情况。根据平稳性及增量独立性，由式(1.2) 可推出:

$$P_n(t+h)=P_n(t)P_0(h)+P_{n-1}(t)P_1(h)+o(h)$$
$$P_n(t+h)=P_n(t)(1-\lambda h+o(h))+P_{n-1}(t)(\lambda h+o(h))+o(h)$$

仿照 n=0 时的方法，移项两边同除以 h，再令 $h\to 0$ 时取极限，有

$$\begin{cases} P_n'(t)=-\lambda P_n(t)+\lambda P_{n-1}(t) & n\geq 1 \\ P_n(0)=P(N(0)=n)=0 & n\geq 1 \end{cases} \tag{1.4}$$

用生成函数的方法来求解上面的微分方程组。令

$$P(t,x)\triangleq\sum_{n=0}^{\infty}P_n(t)\cdot x^n \tag{1.5}$$

将式(1.4) 两端同乘以 x^n，且与式(1.3) 联立，得

$$\begin{cases} P_0'(t)=-\lambda P_0(t) \\ P_n'(t)x^n=-\lambda P_n(t)x^n+\lambda P_{n-1}(t)x^n \end{cases} \tag{1.6}$$

将(1.6) 式两边求和，有

$$\sum_{n=0}^{\infty} P_n'(t)x^n = -\lambda\sum_{n=0}^{\infty} P_n(t)x^n + \lambda x\sum_{n=1}^{\infty} P_{n-1}(t)x^{n-1}$$

即 $\dfrac{\partial P(t,x)}{\partial t} = -\lambda(1-x)P(t,x)$，分离变量后积分，得

$$\ln P(t,x) - \ln P(0,x) = -\lambda(1-x)\,t$$

因
$$P(0,x) = \sum_{n=0}^{\infty} P_n(0)x^n = P_0(0) = 1$$

得
$$P(t,x) = e^{-\lambda t(1-x)} = \sum_{n=0}^{\infty} \frac{(\lambda t)^n e^{-\lambda t}}{n!} x^n$$

将上式与(1.5) 式相比较，可得(1.1) 式成立。

由定理 1.1 可知 $N(t)\sim P_O(\lambda t)$，再由第四章例 1.4 得出 $EN(t)=\lambda t$，$EN(1)=\lambda$。可见λ表示单位时间上到达的平均信号数。

由定理 1.1 容易证明以下

定理 1.2 若信号流$\{N(t), t\geq 0\}$满足：*i*）$N(0)=0$；*ii*）增量独立性；*iii*）$\forall s,t\geq 0$, $n\in \mathrm{N}$，$P(N(s+t)-N(s)) = \dfrac{(\lambda t)^n}{n!}e^{-\lambda t}$。则$\{N(t), t\geq 0\}$是参数为$\lambda$的泊松信号流。

这样定理 1.2 中的 *i*）、*ii*）、*iii*）可作为泊松信号流的另一定义。

§2 用相继到达的时间间隔刻划泊松流

在介绍泊松信号流的主要性质之前，先引入几个具有客观背景的变量。记 $S_0=0, S_n=$“第 n 个信号到达的时刻”，$X_n=S_n-S_{n-1}=$“相继到达的第 $n-1$ 个信号与第 n 个信号之间的时间间隔”，即

$$S_0=0,\ S_n = \sum_{k=1}^{n} X_k$$

当 $N(t)$ 表示$[0,t]$上到达的信号数时，即$\{N(t), t\geq 0\}$ 为计数过程时，则上述 S_1 及 S_n 可确切用 $N(t)$来定义如下

$$\begin{cases} S_0 = 0, S_1 \triangleq \inf\{t; t>0, N(t)\neq 0\} \\ S_n \triangleq \inf\{t; t>S_{n-1}, N(t)\neq n-1\} \quad n>1 \\ X_n = S_n - S_{n-1} \quad n\geq 1 \end{cases} \tag{2.1}$$

其中 inf{ }表示集合{ }的下确界，即集合{ }的最大下界，例如：$\inf\{\frac{1}{n}, n \geq 1\} = 0$ 。

反之，若已给$\{X_n, n\geq1\}$为相继到达信号的时间间隔。又令 $S_0=0$，$S_n = \sum_{k=1}^{n} X_k$ ，也可用 S_0，S_1，…，S_n，…定义 $N(t)$如下：对$\forall t>0$

$$\begin{cases} N(0) \hat{=} 0 \\ N(t) \hat{=} \max\{n; n \geq 0, S_n \leq t\} = \sum_{n=1}^{\infty} I_{(S_n \leq t)} \end{cases} \tag{2.2}$$

由 S_n 与 $N(t)$的定义及关系，我们可以得出当 $t\geq0$ 时有：$(S_n \leq t) \subset (N(t) \geq n)$，且 $(S_n \leq t) \supset (N(t) \geq n)$ 故

$$\begin{aligned} (S_n \leq t) &= (N(t) \geq n) \\ (N(t) = n) = (S_n \leq t < S_{n+1}) &= (S_n \leq t) - (S_{n+1} \leq t) \end{aligned} \tag{2.3}$$

由于 $S_n>0$ 是非负的 *r.v.*，只需考虑 $t\geq0$ 时的分布函数，记 $F_n(t)=P(S_n\leq t)$。由(2.2)，(2.3)及$(S_{n+1}\leq t)\subset (S_n\leq t)$，可得

$$F_n(t)=P(S_n\leq t)=P(N(t)\geq n) \tag{2.4}$$

$$P(N(t)=n)=P(S_n\leq t) - P(S_{n+1}\leq t)=F_n(t) - F_{n+1}(t) \tag{2.5}$$

以上关系很重要，且对任一计数过程$\{N(t),t\geq0\}$与其到达时刻$\{S_n,n\geq1\}$均是如此，以下考察当$\{N(t),t\geq0\}$为简单 *Possion* 流时，它们之间的具体关系。

先看 $n=1$ 的情形，若$\{N(t),t\geq0\}$是参数为λ 的 *p.p.*时， $P(X_1\leq t)=P(S_1\leq t)=P(N(t)\geq1)=1-e^{-\lambda t}$。可知 $X\sim E(\lambda)$是指数分布（参数为λ）。由此可猜想$\{X_n,n\geq 1\}$的规律特性是什么?为此，讨论$\{S_n,n\geq1\}$的特性。

引理 2.1 若$\{N(t),t\geq0\}$是参数为λ的 *p.p.*，则 S_n 的分布函数为：

$$F_n(t) = P(S_n \leq t) = 1 - e^{-\lambda t} \cdot (\sum_{k=0}^{n-1} \frac{(\lambda t)^k}{k!}) \tag{2.6}$$

证明：由定理 1.1 的(1.1)式代入(2.4) 式，即得(2.6)。

引理 2.2 若$\{N(t),t\geq0\}$是参数为λ的 *p.p.*,则：$(S_1, S_2, \cdots S_n)$ 的联合 *p.d.f.* 为

$$f_{(S_1,S_2\cdots,S_n)}(y_1, y_2, \cdots, y_n) = \lambda^n e^{-\lambda y_n} \cdot I_{(0<y_1<y_2<\cdots<y_n)} \tag{2.7}$$

证明：为避免繁琐，下面只对 $n=2$ 情形给出证明（$n\geq3$ 可类似推出）。$n=2$ 时，由于 $0\leq S_1\leq S_2$,只需考虑取 $0\leq y_1<y_2$ 及充分小的 $h_1,h_2>0$ 满足：$0 < y_1 < y_1 + h_1 < y_2 < y_2 + h_2$ 由下面易知

$$\{y_1 < S_1 \leq y_1 + h_1, y_2 < S_2 \leq y_2 + h_2\} =$$

$$\{N(y_1)=0, N(y_1+h_1)-N(y_1)=1, N(y_2)-N(y_1+h_1)=0, N(y_2+h_2)-N(y_2)=1\}$$

$$+\{N(y_1)=0, N(y_1+h_1)-N(y_1)=1, N(y_2)-N(y_1+h_1)=0, N(y_2+h_2)-N(y_2)\geq 2\}$$

再由$\{N(t),t\geq 0\}$增量独立性及定理 1.1 的 (1.1)式得

$$P(y_1<S_1\leq y_1+h_1, y_2<S_2\leq y_2+h_2)=(e^{-\lambda y_1})\cdot(\lambda h_1 e^{-\lambda h_1})\cdot(e^{-\lambda(y_2-y_1-h_1)})\cdot(\lambda h_2 e^{-\lambda h_2})+o(h_1 h_2)$$

$$=\lambda^2\cdot e^{-\lambda y_2}\cdot h_1 h_2\cdot e^{-\lambda y_2}+o(h_1 h_2)=\lambda^2\cdot e^{-\lambda y_2}\cdot h_1 h_2(1-\lambda h_2+o(h_2))+o(h_1 h_2) \text{ 及}$$

$$f_{(S_1,S_2)}(y_1,y_2)=\underset{\substack{h_1\to 0\\ h_2\to 0}}{Lim}\frac{P(y_1<S_1\leq y_1+h_1, y_2<S_2\leq y_2+h_2)}{h_1 h_2}$$

故

$$f_{(S_1,S_2)}(y_1,y_2)=\lambda^2 e^{-\lambda y_2}\cdot I_{(0<y_1<y_2)}$$

对 $n\geq 3$ ，可类似证明。

再分析相邻事件到达的时间间隔 $\{X_n,n\geq 1\}$的特性。

引理 2.3 若$\{N(t),t\geq 0\}$是参数为 λ的简单 *p.p.*, 则相邻两个事件发生的时间间隔$\{X_n,n\geq 1\}$彼此独立，且均服从参数同为λ的指数分布，即$\{X_n,n\geq 1\}$ *i.i.d* ,$X_n\sim E(\lambda)$。

证明：证明大致思路与步骤如下：由$(S_1,S_2,\cdots,S_n)$的 *j.p.d.f.*通过变换 $X_n=S_n-S_{n-1}$ 求$(X_1,X_2,\cdots,X_n)$的 *j.p.d.f.*。再证 $X_i(i=1,2,\ldots,n)$ *i.i.d.*且 $X_n\sim E(\lambda)$。由

$$\begin{cases}X_1=S_1\\X_2=S_2-S_1\\\cdots\\X_n=S_n-S_{n-1}\end{cases}\text{；令：}\begin{cases}x_1=y_1\\x_2=y_2-y_1\\\cdots\\x_n=y_n-y_{n-1}\end{cases}\Rightarrow\begin{cases}y_1=x_1\\y_2=x_1+x_2\\\cdots\\y_n=x_1+x_2+\cdots+x_n\end{cases}$$

求得变换的雅可比行列式为

$$J=\begin{vmatrix}1&0&0&\cdots&0\\1&1&0&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\1&1&1&1&1\end{vmatrix}=1$$

于是$(X_1,X_2,\cdots,X_n)$的 *j.p.d.f.*为:

$$f_{(X_1,X_2\cdots,X_n)}(x_1,x_2,\cdots,x_n)=\lambda^n e^{-\lambda(x_1+x_2+\cdots+x_n)}\cdot I_{(x_i\geq 0,\,1\leq i\leq n)}$$

为求 X_k 的 *p.d.f.*只需将上式对 $x_1,\cdots,x_{k-1},x_{k+1},\cdots,x_n$ 变元积分。结果为：

$$f_k(x_k)=\lambda e^{-\lambda x_k}I_{(x_k\ge 0)}$$

可知，对 $\forall x_k\ge 0, 1\le k\le n$ 有 $f_{(X_1,X_2\cdots,X_n)}(x_1,x_2,\cdots,x_n)=\prod_{k=1}^{n}f_k(x_k)$，即 $\{X_n,n\ge 1\}$ *i.i.d.* 且 $X_n\sim E(\lambda)$。引理得证。

以上说明，相邻时间间隔 $\{X_n,n\ge 1\}$ 独立同指数分布是简单 *Poisson* 流的必要条件，那么它是否是充分条件？换句话说，它是 *p.p.* 所特有的性质？回答是肯定的。

引理 2.4 设 $\{X_n,n\ge 1\}$ 表示相继发生的事件（信号到达）的时间间隔，若它们独立，服从参数为λ的指数分布，则由(2.2)式定义的 $N(t)$ 服从 *Poisson* 分布，即 $\forall\ t\ge 0$，有:

$$P(N(t)=n)=\frac{(\lambda t)^n}{n!}e^{-\lambda t}\qquad (n=0,1,2\cdots)$$

证明： 为求 $N(t)$ 的分布，先求出 S_n 的分布 $F_n(t)\triangleq P(S_n\le t)$，用归纳法。

$$F_n(t)=1-e^{-\lambda t}\cdot(\sum_{k=0}^{n-1}\frac{(\lambda t)^k}{k!})\quad t\ge 0 \tag{2.8}$$

当 $n=1$ 时，易知 $F_1(t)=P(S_1\le t)=P(X_1\le t)=(1-e^{-\lambda t})I_{(t\ge 0)}$

设当 $n=k$ 时，（2.8）成立，须证 $n=k+1$ 时，由 $S_{n+1}=S_n+X_{n+1}$，利用卷积公式

$F_{n+1}(t)=\int_0^t F_n(t-u)dF_1(u)$，有

$$F_{k+1}(t)=\int_0^t[1-e^{-\lambda(t-u)}\cdot(\sum_{i=0}^{k-1}\frac{(\lambda(t-u))^i}{i!})]\cdot\lambda e^{-\lambda u}du=1-e^{-\lambda t}-e^{-\lambda t}\cdot\sum_{i=0}^{k-1}\frac{(\lambda t)^{i+1}}{(i+1)!}$$

$$=1-e^{-\lambda t}\cdot\sum_{i=0}^{k}\frac{(\lambda t)^i}{i!}$$

从而 (2.8)得证。再由(2.8)式，代入(2.5)式，即证结论成立。

由引理 2.4 易得如下推论：

推论： 若 $\{X_n,n\ge 1\}$ *i.i.d*，且 $X_n\sim E(\lambda)$。则由(2.2)式定义的 $\{N(t),t\ge 0\}$ 具有普通性，即 $\forall t\ge 0$ 及充分小的 $h>0$，有

$$P(N(h)=1)=\lambda h+o(h)\qquad\qquad P(N(h)\ge 2)=o(h)$$

引理 2.5 若 $\{X_n,n\ge 1\}$ *i.i.d*，$X_n\sim E(\lambda)$，则由(2.2)式定义的 $\{N(t),t\ge 0\}$ 具有增量平稳性，即 $\forall s,t\ge 0$，$P(N(s+t)-N(s)=n)=P(N(t)=n)\ \forall n\in N$。

证明： 这里只写出对 $n\ge 1$ 的证明如下:利用全概率公式及（2.3）式，有

$$P(N(s+t)-N(s)=n)=\sum_{k=0}^{\infty}P(N(s)=k,N(s+t)=k+n)$$

$$= \sum_{k=0}^{\infty} P(S_k \le s < S_{k+1} \le S_{k+n} \le s + t < S_{k+n+1})$$

$$= P(s < S_1 \le S_n \le s + t < S_{n+1}) +$$

$$\sum_{k=1}^{\infty} \int_0^s P(S_k \le s < S_k + X_{k+1} \le S_k + \sum_{k+1}^{k+n} X_i \le s + t \le S_k + \sum_{k+1}^{k+n+1} X_i \mid S_k = u) dP(S_k \le u)$$

$$= P(s < S_1 \le S_n < s + t < S_{n+1}) + \sum_{k=1}^{\infty} \int_0^s P(s - u < S_1 \le S_n \le s + t - u < S_{n+1}) dP(S_k \le u)$$

又

$$P(s < S_1 \le S_n \le s + t < S_{n+1}) = \int_s^{s+t} P(\sum_2^n X_i \le s + t - v < \sum_2^{n+1} X_i \mid X_1 = v) \lambda e^{-\lambda v} dv$$

$$= \int_s^{s+t} P(S_{n-1} \le s + t - v < S_n) \lambda e^{-\lambda v} dv = \int_s^{s+t} \frac{[\lambda(s + t - v)]^{n-1}}{(n-1)!} e^{-\lambda(s+t-v)} \lambda e^{-\lambda v} dv$$

$$= \frac{(\lambda t)^n}{n!} e^{-\lambda(s+t)} \qquad (*)$$

类似地有 $P(s-u<S_1\le S_n\le s+t-u<S_{n+1})=\frac{(\lambda t)^n}{n!}e^{-\lambda(s+t-u)}$，代入（*）式得

$$P(N(s+t) - N(s) = n) = \frac{(\lambda t)^n}{n!} e^{-\lambda(s+t)} + \int_0^s \frac{(\lambda t)^n}{n!} e^{-\lambda(s+t-u)} \lambda du$$

$$= \frac{(\lambda t)^n}{n!} e^{-\lambda(s+t)} + \frac{(\lambda t)^n}{n!} e^{-\lambda(s+t)} (e^{\lambda s} - 1) = \frac{(\lambda t)^n}{n!} e^{-\lambda t} = P(N(t) = n)$$

注：对 $n=0$ 情形类似证明，留作读者练习。

引理 2.6 若$\{X_n,n\ge1\}$ $i.i.d$,$X_n\sim E(\lambda)$，则由(2.2)式定义的$\{N(t),t\ge0\}$具有增量独立性，即 $\forall 0 \le t_1 < t_2 < \cdots < t_n$， $N(t_1) - N(0), N(t_2) - N(t_1), \cdots, N(t_n) - N(t_{n-1})$ 相互独立。

证明：为节省篇幅，下面只证对$\forall s,t>0$，$N(s)$与 $N(s+t)-N(s)$ 独立，即证：$\forall m,n\ge0$ 有：$P(N(s)=m,N(s+t)-N(s)=n)=P(N(s)=m)P(N(s+t)-N(s)=n)$

只写出 $m\ge1$ 与 $n\ge1$ 的情形，其它可类似证之。

$$P(N(s) = m, N(s+t) - N(s) = n) = P(S_m \le s < S_{m+1} \le S_{m+n} \le s + t < S_{m+n+1})$$

$$= \int_0^s P(s - u < X_{m+1} \le \sum_{m+1}^{m+n} X_i \le s + t - u \le \sum_{m+1}^{m+n+1} X_i \mid S_m = u) dP(S_m \le u)$$

$$= \int_0^s P(s - u < S_1 \le S_n < s - u + t < S_{n+1}) dP(S_m \le u)$$

$$= \int_0^s \frac{(\lambda t)^n}{n!} e^{-\lambda(s-u+t)} \lambda \frac{(\lambda t)^{m-1}}{(m-1)!} e^{-\lambda u} du = \frac{(\lambda s)^m}{m!} e^{-\lambda s} \frac{(\lambda t)^n}{n!} e^{-\lambda t}$$

$$= P(N(s) = m) P(N(s+t) - N(s) = n)$$

综合引理 2.2 ~2.6，我们有以下重要定理：

定理 2.1 计数过程$\{N(t),t\geq 0\}$是简单 *Possion* 流的充要条件是其信号相继到达的时间间隔$\{X_n,n\geq 1\}$ *i.i.d.*,且$X_n\sim E(\lambda)$。

以上结果揭示了简单 *Possion* 流与服从指数分布且相互独立的时间间隔的内在联系，这不仅在理论上有价值，而且在应用上很有意义。例如，要检验一个计数过程是否是 *p.p.*问题，可以转化为检验$\{X_n,n\geq 1\}$是否 *i.i.d.* 及$X_n\sim E(\lambda)$。同时，要在计算机上模拟一简单 *Possion* 流 $\{N(t),t\geq 0\}$。可通过以下步骤构造出:

i) 取一串$\{U_n,n\geq 1\}$ *i.i.d.* $U_n\sim U(0,1)$;

ii) 令$X_n=-\lambda^{-1}\ln U_n \quad (\lambda>0)$ 易知，$\{X_n,n\geq 1\}$ *i.i.d.*,且$X_n\sim E(\lambda)$;

iii) 令$S_0=0, S_n=\sum_1^n X_k$;

iv) $\forall t\geq 0$ 定义 $N(t)=\sum_{n=1}^{\infty} I_{(S_n\leq t)}=\sup\{n: n\geq 0, S_n\leq t\}$

则$\{N(t),t\geq 0\}$为 *p.p.*。

§3 相继到达时刻的条件分布

考虑在$N(t)=n$条件下，$(S_1,S_2,\cdots S_n)$的条件分布。先看在$N(t)=1$下，S_1 的 *c.p.d.f.* ，由于$S_1\geq 0$ ，同时在$N(t)=1$下，$S_1\leq t$, 故只需考察$0\leq x\leq t$ 的情形。当$0\leq x\leq t$时，有：

$$P(S_1\leq x\mid N(t)=1)=\frac{P(S_1\leq x,N(t)=1)}{P(N(t)=1)}=\frac{P(N(x)=1,N(t)-N(x)=0)}{P(N(t)=1)}=\frac{x}{t}$$

$$f_{S_1|N(t)=1}(x)=\frac{1}{t}\cdot I_{(0\leq x\leq t)} \tag{3.1}$$

可见$S_1=X_1$ 在$N(t)=1$下的 *c.p.d.f.* 是$[0,t]$上的均匀分布，一般地，我们有

定理 3.1 设$\{N(t),t\geq 0\}$为泊松信号流，则在$N(t)=n$ $(n\geq 1)$下$(S_1,S_2,\cdots,S_n)$的 *c.p.d.f.*为

$$f(x_1,x_2,\cdots,x_n)=\frac{n!}{t^n}\cdot I_{(0<x_1<x_2<\cdots<x_n<t)} \tag{3.2}$$

证明：为节省篇幅仅证$n=2$。因在$N(t)=2$下, $0\leq S_1\leq S_2\leq t$,故只考虑$0<x_1<x_2<t$，取充分小$h>0$ 使$0<x_1<x_1+h<x_2<x_2+h<t$,于是

$$P(x_1<S_1\leq x_1+h,x_2<S_2\leq x_2+h\mid N(t)=2)$$

$$=\frac{P\{N(x_1)=0,N(x_1+h)-N(x_1)=1,N(x_2)-N(x_1+h)=0,N(x_2+h)-N(x_2)=1,N(t)-N(x_2+h)=0\}}{P(N(t)=2)}$$

$$= e^{-\lambda x_1}(\lambda h e^{-\lambda h})e^{-\lambda(x_2-x_1-h)}(\lambda h e^{-\lambda h})e^{-\lambda(t-x_2-h)}\Big/(\frac{(\lambda t)^2}{2!}e^{-\lambda t}) = \frac{2!}{t^2}h_1 h_2$$，得

$f_{(S_1,S_2|N(t)=2)}(x_1,x_2) = \frac{2!}{t^2} I_{(0<x_1<x_2<t)}$。$n$≥3 的情况可类似证得。

本定理说明在 $N(t)=n$ 的条件下 $(S_1,S_2,\cdots,S_n)$ 的 *c.p.d..f.*与 n 个在[0,t]上相互独立同分布的顺序统计量的 *j.p.d..f.*相同（见第三章命题 7.1）。其逆命题如下：

定理 3.2 设$\{N(t),t\ge0\}$为计数过程（信号流）。$\{X_n,n\ge0\}$为信号相继到达的时间间隔，独立同分布 $F(x) = P(X_n \le x), F(0) = 0$，且 $\forall 0 \le s \le t$，$P(X_1 \le s \mid N(t)=1) = (s/t)\cdot I_{(0\le s\le t)}$，则$\{N(t),t\ge0\}$为泊松信号流。

证明：由题设 $P(X_1 \le s \mid N(s+x)=1) = s/(s+x), P(X_1 \le x \mid N(s+x)=1) = x/(s+x)$，所以

$$P(X_1 \le s \mid N(s+x)=1) + P(X_1 \le x \mid N(s+x)=1) = 1 \qquad x,s>0 \qquad (3.3)$$

又 $P(X_1 \le s \mid N(s+x)=1) = P(X_1 \le s, X_1 \le s+x < X_1+X_2)/P(X_1 \le s+x < X_1+X_2)$

由全概率公式可得 $P(X_1 \le s, X_1 \le s+x < X_1+X_2) = \int_0^s (1-F(s+x-u))dF(u)$。

同理 $P(X_1 \le s+x < X_1+X_2) = \int_0^{x+s}(1-F(s+x-u))dF(u)$。

由（3.3）式可得

$\int_0^x (1-F(s+x-u))dF(u) + \int_0^s (1-F(s+x-u))dF(u) = \int_0^{x+s}(1-F(s+x-u))dF(u)$，所以

$$\int_0^s (1-F(s+x-u))dF(u) = \int_x^{x+s}(1-F(s+x-u))dF(u) \quad x,s>0。$$

令 $G(x)=1-F(x)$，则上式可化为

$$G(s+x)= G(s)G(x) \qquad s,x\ge0 \qquad (3.4)$$

由 $G(x)$单调不增、右连续，则存在右导数 $G'_x(x)$；又 $G(0)=1$，所以 $G'_s(x+s) = G'_s(s)G(x)$，

又 $G'_x(x+s) = G'_s(x+s)$ 令 $s=0$，则 $G'_x(x) = G'_s(0)G(x)$。令 $\lambda = -G'_s(0) \ge 0$，则

$G(x) = e^{-\lambda x}$，$x \ge 0$。由 $F(x)$是分布函数，不能恒等于零，可得 $\lambda > 0$，从而 $F(x)$服从指数分布。再由定理 2.1 即证。

注：用以上结果检验泊松信号流时，不必知道参数λ。

例 3.1 求 $S(t) \triangleq \sum_{k=1}^{N(t)}(t-S_k)$ 的期望 $ES(t)$，其中$\{N(t),t\ge0\}$是参数为λ的 *p.p.*。S_n 为

其第 n 个事件发生时刻。此例的背景之一是：设 t 为火车离站时刻，S_k 为第 k 个顾客到站时刻,那么 $S(t)$为在$(0,t)$ 上到站乘客的候车时间之和。

解：由条件期望性质，有

$$ES(t)=E\{E[\sum_{k=1}^{N(t)}(t-S_k)\mid N(t)]\}$$

为此先求

$$E[S(t)\mid N(t)=n]=E[\sum_{k=1}^{N(t)}(t-S_k)\mid N(t)=n]=E[\sum_{k=1}^{n}(t-S_k)\mid N(t)=n]$$

$$=nt-E[\sum_{k=1}^{n}S_k\mid N(t)=n]=nt-E[\sum_{k=1}^{n}U_{(k)}]=nt-E[\sum_{k=1`}^{n}U_k]=nt-\sum_{k=1}^{n}EU_k=\frac{nt}{2}$$，故

$$ES(t)=E\{E[S(t)\mid N(t)]\}=\sum_{n=0}^{\infty}E[S(t)\mid N(t)=n]P(N(t)=n)=\frac{\lambda t^2}{2}。$$

§4 剩余寿命与年龄

本节再从不同角度刻划 *Possion* 流的特性。设 $N(t)$表示在$[0,t]$ 上事件发生的个数，S_n表示第 n 个事件发生的时刻，那么，$S_{N(t)}$表示在 t 时刻前最后一个事件发生的时刻。$S_{N(t)+1}$表示 t 时刻后首次事件发生的时刻。注意这里 $S_{N(t)}$与 $S_{N(t)+1}$ 的下标 $N(t)$与 $N(t)+1$ 是随机变量。令

$$W(t)=S_{N(t)+1}-t \tag{4.1}$$

$$V(t)=t-S_{N(t)} \tag{4.2}$$

$W(t)$与 $V(t)$ 有许多具体背景，例如，设一零件在 t=0 时开始工作，若它失效，立即更换，一个更新的零件又开始工作，如此重复下去，记 S_n 为第 n 次更换时刻，记 $X_n=S_n-S_{n-1}$ 表示第 n 个零件的工作寿命。此时 $W(t)$表示观察者在时刻 t 所观察的正在工作的零件的剩余寿命；$V(t)$ 表示正在工作的零件已经工作的时间，称为 年龄。还可以有别的解释：如 S_n表示第 n 辆汽车到站的时刻，某一乘客到站的时刻为 t, 则 $W(t)$ 表示该乘客的等待时间。若 S_n 表示第 n 次发生地震的时刻，则 $S_{N(t)+1}$ 表 t 时刻以后首次地震的时刻，$W(t)$ 表示 t 时刻后直到首次地震之间的剩余时间，等等。所以，通常称 $W(t)$为剩余寿命；$V(t)$ 为年龄，显然，研究 $W(t)$, $V(t)$的特性很有意义。

定理 4.1 若$\{N(t),t\geq 0\}$ 是参数为λ的简单 *Poisson* 信号流，则:

ⅰ) $W(t)$与$\{X_n,n\geq 1\}$ 同为指数分布；

ⅱ) $V(t)$是“截尾”的指数分布，即:

$$P(V(t)\leq x)=\begin{cases}1-e^{-\lambda x} & 0\leq x<t\\ 1 & x\geq t\end{cases} \tag{4.3}$$

证明： 由$\{W(t)>x\}=\{N(t+x)-N(t)=0\}$，及

$t>x$时 $\{V(t)>x\}=\{N(t)-N(t-x)=0\}$， $t\le x$时 $\{V(t)>x\}=\phi$，即得所要结论。

推论 若$\{N(t),t\ge0\}$是参数为λ的简单 *p.p.*，则 $EW(t)=1/\lambda$， $EV(t)=(1-e^{-\lambda t})/\lambda$。

证明 由定理 1.4 易得，请读者自己完成。

问题：$\forall t>0, W(t), V(t)$ 是否独立？请有兴趣的读者证明你的猜想。

还可用 $W(t)$与 X_n 的关系来刻划泊松信号流。

定理 4.2 设$\{X_n, n\ge1\}$ *i.i.d.*， $X_n\ge0$。$\forall t\ge0, N(t)$ 由(2.2)式定义，$W(t)$由(4.1)式定义，若 $W(t)$与 X_n 独立同分布 $F(x)$，且 $F(0)=0$，则$\{N(t),t\ge0\}$是泊松信号流。

证明：梗概如下：

设 $G(x)=1-F(x)$，由 $W(t)$的定义及全概率公式可证：$\forall x,t\ge0$，有

$$P(W(t)>x)=1-F(x+t)+\int_0^t P(W(t-u)>x)dF(u) \tag{4.4}$$

由此可得 $G(t+x)=G(t)G(x)\quad t,x\ge0$。类似与定理 3.2 最后部分的证明可证结论。

§5 泊松流的若干推广

在前面的讨论中，我们看到 *Poisson* 信号流有许多良好的独特性质，但现实问题中并不都能满足 *p.p.*定义中的所有条件。因此，有必要讨论定义中某些条件放宽或减弱后的情形，下面提一下它的一些推广。

1 非时齐 *Poisson* 流

把 *p.p.* 定义中将增量平稳性放宽的情况，即去掉λ是常数的限制，即得到非时齐 *Poisson* 流。

定义 5.1 计数过程$\{N(t),t\ge0\}$称为具有强度函数$\{\lambda(t),t\ge0\}$的非时齐 *p.p.*若它满足

（1）$N(0)=0$;

（2）$\{N(t),t\ge0\}$具有增量独立性；

（3）$\forall t\ge0$，充分小的 $h>0$, $P(N(t+h)-N(t)=1)=\lambda(t)h+o(h)$, $P(N(t+h)-N(t)\ge2)=o(h)$,其中$\{\lambda(t),t\ge0\}$称为强度函数，$\lambda(t)\ge0$。

定理 5.1 若$\{N(t),t\ge0\}$是强度函数$\{\lambda(t),t\ge0\}$的非时齐 *p.p.*，令 $m(t)=\int_0^t\lambda(u)du$，则$\forall s,t\ge0$ 有

$$P(N(s+t)-N(s)=n)=\frac{[m(s+t)-m(s)]^n}{n!}\exp\{-[m(s+t)-m(s)]\}, n\in\mathrm{N} \tag{5.1}$$

证明：证法与定理一的证法完全类似，留作练习。

可以证明，通过适当的变换可将非时齐的 *p.p.*转化为时齐的 *p.p.*，反之亦然。

2.复合 *Poisson* 流

定义 5.2 设$\{Y_k, k\geq 1\}$,*i.i.d.*，$\{N(t),t\geq 0\}$为简单 *p.p.*且$\{N(t),t\geq 0\}$与$\{Y_k, k\geq 1\}$独立，记

$$X(t)=\sum_{k=1}^{N(t)}Y_k, \tag{5.2}$$

称$\{X(t),t\geq 0\}$为复合 *p.p.*。

特殊情况，若$\{\rho_k,k\geq 1\}$ *i.i.d.*，是取值为正整数的 *r.v.* 序列，且$\{\rho_k,k\geq 1\}$与$\{N(t),t\geq 0\}$独立，记 $X(t)=\sum_{k=1}^{N(t)}\rho_k$，则称$\{X(t),t\geq 0\}$为平稳无后效流。容易理解 $X(t)$可用于描述成批到达的顾客流。

3 条件 *Poisson* 流

定义 5.3 设Λ是一正随机变量，分布函数为 $G(x),x\geq 0$。设$\{N(t),t\geq 0\}$是计数过程，且在给定Λ=λ的条件下，$\{N(t),t\geq 0\}$是一时齐的 *Poisson* 流，即 $\forall s,t\geq 0,\ \lambda\geq 0, n\in N$。有：

$$P(N(s+t)-N(s)=n\mid\Lambda=\lambda)=\frac{(\lambda t)^n}{n!}e^{-\lambda t} \tag{5.3}$$

4 更新过程

由定理 2.1 知，一个计数过程，若他们相邻到达的时间间隔 X_n 是指数分布，则此过程为 *Poisson* 流。现在，我们来考虑 X_n 是一般分布时的情形，这便是更新过程。

定义 5.4 设$\{X_n,n\geq 0\}$*i.i.d*，且 $X_n\geq 0$，分布函数为 $F(x),x\geq 0$，且 $F(0)<1$，令 $S_0=0$，$S_n=\sum_{k=1}^{n}X_k$，对$\forall t\geq 0$,记 $N(t)=sup\{n: n\geq 0,S_n\leq t\}$，称$\{N(t),t\geq 0\}$为更新过程。

显然，更新过程是一计数过程，并有

$$\{N(t)\geq n\}=\{S_n\leq t\}, \tag{5.4}$$

$$\{N(t)=n\}=\{S_n\leq t<S_{n+1}\}=\{S_n\leq t\}-\{S_{n+1}\leq t\} \tag{5.5}$$

请读者考虑：更新过程是否具有增量独立性。

记 $F_n(x)$为 S_n 的分布函数，由 $S_n=\sum_{k=1}^{n}X_k$，易知：

$$F_1(x)=F(x)$$

$$F_n(x)=\int_0^x F_{n-1}(x-u)dF(u)\quad (n\geq 2)$$

即 $F_n(x)$是 $F(x)$的 n 重卷积（简记 $F_n=F_{n-1}*F$）。记 $m(t)=EN(t)$，称 $m(t)$为更新函数。

定理 5.2 $\forall t\ge 0$, 有

$$m(t)=\sum_{n=1}^{\infty}F_n(t). \tag{5.6}$$

证明： $m(t)=\sum_{n=0}^{\infty}nP(N(t)=n)=\sum_{n=0}^{\infty}P(N(t)\ge n)=\sum_{n=0}^{\infty}P(S_n\le t)=\sum_{n=1}^{\infty}F_n(t)$。

推论 若对$\forall t\ge 0, F(t)<1$，则

$$m(t)\le F(t)(1-F(t))^{-1} \tag{5.7}$$

证明：由归纳法可得：$F_n(t)\le (F(t))^n$，再利用(5.6)即得。

定理 5.3 $\forall t\ge 0$, $m(t)$满足下列更新方程

$$m(t)=F(t)+\int_0^t m(t-u)dF(u). \tag{5.8}$$

证明：由(5.6)得 $m(t)=F(t)+\sum_{n=2}^{\infty}F_n(t)$，将 $F_n(x)=\int_0^x F_{n-1}(x-u)dF(u)$ 代入即得(5.8)。

若令 $\widetilde{m}(s)=\int_0^{\infty}e^{-st}dm(t)$， $\widetilde{F}(s)=\int_0^{\infty}e^{-st}dF(t)$, 从(5.8)式易得

$$\widetilde{m}(s)=\frac{\widetilde{F}(s)}{1-\widetilde{F}(s)}\ ，\widetilde{F}(s)=\frac{\widetilde{m}(s)}{1+\widetilde{m}(s)} \tag{5.9}$$

由拉氏变换与其逆变换是一一对应的，可知 $m(t)$与 $F(t)$亦是一一对应的。

令 $\mu=EX_n$， 由 $F(0)<1$ 易证$\mu>0$。记 $N(\infty)=\lim_{t\to\infty}N(t)$ 。

定理 5.4（1） $P(\lim_{n\to\infty}\frac{S_n}{n}=\mu)=1$，（2） $P(N(\infty)=\infty)=1$，（3） $P(\lim_{t\to\infty}\frac{N(t)}{t}=\frac{1}{\mu})=1$。

证明：（1）由强大数定律即得。

（2）注意到 $\{N(\infty)<\infty\}=\bigcup_{n=1}^{\infty}(S_n=\infty)=\bigcup_{n=1}^{\infty}(X_n=\infty)$ 即得。

（3）从略（留给有兴趣的读者作为练习来完成）。

更新过程的最初物理原型是零件的连续更换，一个零件在零时刻开始工作，在 X_1 时刻失效，然后马上被第二个更换，一般地，第 n 个零件在 $\sum_{i=1}^{n}X_i$ 时失效，随之马上换一新零件。通常假定各零件寿命是独立同分布的，即 $P(X_n\le x)=F(x)$，显然在这一更换过程中 $N(t)$表示在$[0,t]$的更新数目。现在，更新过程在生物遗传，排水系统，可靠性工程，人口增长，经济管理等领域有着广泛应用。

练习题

6.1 假设在某一公路上，同方向行驶的汽车之间的距离是具有均值为 100 米的指数分布且相互独立。问：在 5 公里的一段路上有 50—60 辆汽车的概率是多少？

6.2 在某一公路上，可以假设运输流为强度等于每分钟 30 辆汽车的简单 *p.p.*，试求出 n 辆汽车（一辆接一辆）通过观察岗需要时间多于 t 秒的概率。

6.3 设 $\{N(t), t \geq 0\}$ 是参数为 λ 的 *Poisson* 流，$S_0 = 0, S_n$ 为第 n 个信号（顾客）到达时刻。（1）下列事件是何关系：$(N(t) < n)$ 与 $(S_n > t)$；$(N(t) \leq n)$ 与 $(S_n \geq t)$；$(N(t) > n)$ 与 $(S_n < t)$；（2）$\forall s, t > 0$，求 $E[N(s)N(t)], P(N(t) = k \mid N(s) = n)$ 及 $E[N(t) \mid N(s)]$ 的分布律；（3）求 E（$S_1 \mid N$（t）$= n$）及 E（$S_1 \mid N$（t））的分布律；（4）求 E（$S_k \big| N(t) = n$）；（5）试问 $W(t) = S_{N(t)+1} - t$ 与 $V(t) = t - S_{N(t)}$ 是否独立？试证之；（6）求 $\delta(t) = S_{N(t)+1} - S_{N(t)}$ 的 $p.d.f.$。

6.4 设 $\{N(t), t \geq 0\}$ 是参数为 λ 的时齐 *Poisson* 过程，$\{Y_k, k \geq 1\}, i.i.d.$ 且与 $\{N(t), t \geq 0\}$ 独立，$EY_1 = \mu, DY_1 = \sigma^2$，令 $X(t) = \sum_{k=1}^{N(t)} Y_k$，求：$EX(t), DX(t)$。

6.5 设一系统在[0，t]内受冲击的次数 $\{N(t), t \geq 0\}$ 是参数为 λ 的时齐 *Poisson* 过程，第 k 次受冲击的即时损失为 $D_k, \{D_k, k \geq 1\} i.i.d.$ 并与 $\{N(t), t \geq 0\}$ 独立，且损失随时间呈指数衰减。$t=0$ 的损失为 D，经 t 时刻损失为 $De^{-\alpha t}$ 设损失可加，t 时刻的总损失为 $\xi(t) = \sum_{k=1}^{N(t)} D_k e^{-\alpha(t - S_k)}$，试求 $E\xi(t), (\alpha > 0$ 为常数)。

6.6 设 $\{N(t), t \geq 0\}$ 是参数为 λ 的 *Poisson* 流，$S_0 = 0, S_1, S_2, \cdots, S_n$ 为相继事件发生时刻。

(1) 给定 $N(t) = n$ 下，试问 $S_1, S_2 - S_1, \cdots, S_n - S_{n-1}$ 是否条件独立？同分布？试证明之；

(2) 求 ES_1 及 $E(S_k \mid N(t)=n), n\ge 0, k=1,2,\cdots$；（3）求 (S_1,S_2) 在 $N(t)=1$ 下的条件 $p.d.f.$；（4）求 (S_2,S_5) 的联合 $p.d.f.$。

6.7 设 $\{N(t),t\ge 0\}$ 为 *Poisson* 过程, 参数为 λ,证明：

（1）任给 $0\le s\le t$, 有 $P\{N(s)\le N(t)\}=1$；

（2）任给 $0\le s\le t, \varepsilon>0$, 有 $\lim_{t\to s} P\{N(t)-N(s)>\varepsilon\}=0$。

6.8 设 U_k 是 *i.i.d.* $U_k \sim (0,1)$， $X_k=-\lambda^{-1}\ln U_k, (\lambda>0)$， 求 (1) X_k 的分布；(2) $S_n=\sum_{k=1}^n X_k$ 的 $p.d.f.$；(3) $Z_n=2\lambda S_n$ 的 $p.d.f.$, 并与 $\chi^2(2n)$ 的 $p.d.f.$比较。

6.9 对固定 t>0，求证：$\forall\varepsilon>0, \lim\limits_{n\to\infty} P\{|\frac{1}{n}\sum_{k=1}^n S_k-\frac{t}{2}|<\varepsilon \big| N(t)=n\}=1$。其中 $N(t)$为泊松信号流，S_n为第 n 个信号到达时刻。

6.10 考虑一更新过程，如果 $P(X_n=1)=1/3, P(X_n=2)=2/3$，计算：$P(N(1)=k)$, $P(N(2)=k), P(N(3)=k)$。

6.11 设 $\{X_n, n\ge 1\}$ $i.i.d.$, X 的 $p.d.f.$为 $f(x)=\rho e^{-\rho(x-\delta)} I_{(x>\delta)}$, 其中 $\delta>0$ 给定， 求更新过程中的概率 $P(N(t)\ge k)$。

6.12 设 X_n 的概率密度 $f(x)=\lambda^2 x e^{-\lambda x}, x\ge 0$, 求相应的更新函数 $m(t)$。

6.13 设 $\{N_i(t),t\ge 0\}$ 是参数分别为 λ_i 的时齐 *Poisson* 过程，且相互独立(i=1,2)。$S_0^{(i)}=0, S_n^{(i)}$ 为第 i 个过程第 n 个事件发生的时刻。

（1）令 $N(t)=N_1(t)+N_2(t)$，证明 $\{N(t),t\ge 0\}$ 是参数为 $\lambda_1+\lambda_2$ 的时齐 *Poisson* 过程；

（2）求 $S_{N_2(t)}^{(1)}$ 的 $p.d.f.$。

6.14 设一信号接收器正同时接收两类相互独立且参数为 λ_i 的 *Poisson* 信号流 $\{N_i(t),t\ge 0\}$， $N_i(0)=0$。$S_n^{(i)}$ 表示第 i 类第 n 个信号到达的时刻，i=1,2,n≥1；

(1) $\forall 0<s<t$，求 $E\{N_1(t)+N_2(t)\mid N(s)\}$ 与 $N_1(S_1^{(1)})$ 的分布律；

（2） $E\{S_1^{(1)}\mid N_1(t)+N_2(t)=n\}, n=1,2$ （3）求不加分类的第 2 个信号到达时刻的数学期望。

6.15 X_1, X_2 为相互独立指数分布随机变量 $P(X_i>t)=e^{-\lambda_i t}, t\ge 0, i=1,2.$

令 $N=\begin{cases}1, & X_1<X_2,\\ 2, & X_2\le X_1,\end{cases}$， $U=\min\{X_1, X_2\}$，$W=|X_1-X_2|$。

证明：（1） $P(N=i)=(\lambda_1+\lambda_2)^{-1}\lambda_i, \qquad i=1,2,$

（2) $P(U>t=\exp\{-(\lambda_1+\lambda_2)t\}, \qquad t\ge 0,$

（3) N 与 U 独立，

（4） $P(W>t\mid N=3-i)=e^{-\lambda_i t}, \qquad t\ge 0, \qquad i=1,2,$

（5） U 与 W 独立。

6.16 某台仪器由 A, B 两个系统构成，可能发生的故障有三种类型。发生第 i 类（i=1，2，3）故障的累计次数构成三个相互独立的强度分别为 λ_i 的 *Poisson* 过程。若发生第一类故障，系统 *A* 无法正常运行；若发生第二类故障，系统 *B* 无法正常运行；若发生第三类故障，系统 *A*、*B* 都无法正常运行。以 X_1 和 X_2 分别记 *A*，*B* 两个系统的寿命。

证明：（1） X_1 和 X_2 的联合分布为：

$$P(X_1>s, X_2>t)=\exp\{-\lambda_1 s-\lambda_2 t-\lambda_3\max(s,t)\};$$

（这个分布通常成为二元指数分布）（2） X_1 和 X_2 均服从指数分布。

6.17 设 $N^{(k)}=\{N^{(k)}(t), t\ge 0\}, k\ge 1$，为一列相互独立的 *Poisson* 过程， $N^{(k)}$ 的强度为 λ_k，又 $\lambda=\sum_{k=1}^{\infty}\lambda_k<\infty$。令 $N(t)=\sum_{k=1}^{\infty}kN^{(k)}(t), t\ge 0$，证明：$N=\{N(t), t\ge 0\}$ 为复合 *Poisson* 过程。

第七章 随机游动与离散鞅

自然科学中的大量问题可归结为随机游动问题，例如随机游动模型可以作为布朗运动的初步近似。概率论中的一些经典问题也能引导到随机游动问题上，事实上，随机游动也是第九章马氏链的直观模型和向导。本章只讨论在一维直线整数点上的随机游动及其应用。

§1 简单随机游动

考虑X轴上的一个质点，假定它只能位于整数点，在时刻t=0时，它处于初始位置$X_0=0$。以后每隔单位时间，它总受到一个外力的随机作用，使位置发生变化，即随机地以概率 p 及概率$q=1-p$ 向右或向左移动一个单位，且各次移动相互独立，我们所关心的是质点在时刻$t=n$时的位置，用这种方式描述的质点运动称为一维直线上的简单随机游动。

若令：$Y_i=\begin{cases}1, & \text{质点第}i\text{次向右移动}\\ -1 & \text{质点第}i\text{次向左移动}\end{cases}$，$X_0=0$，$X_n=\sum_{i=1}^{n}Y_i$，则$X_n$表质点（开始从原点出发）第$n$时刻的位置。

关于质点在一维直线上整数点的随机游动，已进行过许多研究，下面我们介绍它的较简单的两个模型。

1 无限制随机游动

在这个模型中，质点可以在整个数轴的整数点上作游动。随机变量序列$\{X_n,\ n\geq 0\}$表示质点在不同时刻的位置，下面分别考虑n=1，2，3$\propto$ 的概率特性。

当n=1时，显然 $X_1=Y_1=\begin{cases}1, & P(X_1=1)=p\\ -1, & P(X_1=-1)=q\end{cases}$，

当n=2时，有

$$X_2=Y_1+Y_2=\begin{cases}2 & P(X_2=2)=P(Y_1=1,Y_2=1)=p^2\\ 0 & P(X_2=0)=P(Y_1=1,Y_2=-1)+P(Y_1=-1,Y_2=1)=2pq\\ -2 & P(X_2=-2)=P(Y_1=-1,Y_2=-1)=q^2\end{cases},$$

类似地，当n=3时，留作练习。

对于任意时刻$t=n$的情况，为了使质点在$t=n$ 时位于k（k 也可以是负整数）必须而

且只需在前n 次移动中向右移动的次数比向左移动的次数多k 。若以x 记它在前n 次游动中向右游动的次数，y记向左游动的次数，则：$x+y=n$ $x-y=k$ 即 $x=\frac{n+k}{2}$，因为x是整数。所以k必须和n具有相同的奇偶性。事件$\{X_n=k\}$发生相当于要求在前 n次游动中有$\frac{n+k}{2}$次向右，$\frac{n-k}{2}$ 次向左，利用二项分布即得

$$P(X_n=k)=C_n^{\frac{n+k}{2}}p^{\frac{n+k}{2}}q^{\frac{n-k}{2}}\text{。} \tag{1.1}$$

而当n与k 奇偶性相反时，概率为0。由(1.1)式，可引出下面一条命题。

命题1.1 若令$S_n=\frac{X_n+n}{2}$ ，则$S_n\sim B(n,p)$，即S_n服从参数为(n,p)的二项分布。

证明：留给读者。

下面来看看一维直线随机游动在竞赛（或赌博）中的应用。

设A、B两队(人)进行某种竞赛(赌博)。每次A胜（赢）的概率为p ，负的概率为q $(p+q=1)$ 规定若A胜，则加1分(得一元)，否则减1分(给B一元)， 且每次之间的胜负相互独立，于是上面定义的X_n 就可以表示进行了n 次竞赛(赌博)后，A的得分（得的钱）。

在上述背景下，我们来求 $E(X_{n+1}|X_0, X_1,\cdots,X_n)$。

分两种情况讨论

1) $p=q=1/2$（对称随机游动）

此时$EY_{n+1}=p-q=0$，并注意到Y_{n+1} 与$X_0, X_1,\cdots,X_n$相互独立。于是$E(X_{n+1}|X_0, X_1,\cdots, X_n)=E(X_n+Y_{n+1}|X_0, X_1,\cdots,X_n)=E(X_n|X_0, X_1,\cdots,X_n)+E(Y_{n+1}|X_0, X_1,\cdots, X_n)=X_n+EY_{n+1}=X_n$。 由此可得

$$E(X_{n+1}|X_0, X_1,\cdots,X_n)=X_n \tag{1.2}$$

上式说明，在每次胜负（输赢）机会相等的情况下，即在公平竞赛（赌博）前提下，在已知前n次结果 $X_0, X_1,\cdots,X_n$的条件下，考虑前$n+1$次平均总得分$E(X_{n+1}|X_0, X_1,\cdots,X_n)$等于前$n$次的总得分。因此，满足(1.2) 式，反映了由于公平竞争（赌博）所导致的序列$\{X_n, n\ge1\}$的一个特性，称具有(1.2)式特性的随机序列为鞅，鞅是近代概率论发展的最为迅速的一个重要分支，它已广泛用于现代随机控制优化理论，离散随机信号的滤波，预测与通讯等工程技术，经济管理以及随机微分方程等基础理论的多个领域。我们将在第三节中对鞅稍加讨论。

2）$p>q$

同理可算得 $E(X_{n+1}|X_0, X_1,\cdots,X_n)=X_n+(p-q)>X_n$。

由于$p>q$，表示每次胜负机会不平等（不公平赌博）。那么，在已知前n次结果$X_0, X_1,\cdots,X_n$ 的条件下，考虑前$n+1$次平均总得分$E(X_{n+1}|X_0, X_1,\cdots,X_n)$ 必然大于前n次的总得分，这样的竞赛（赌博）显然是对其中一方有利，所以是不公平的。

2 求输光的概率

如果参加赌博时A有资本a元，B有资本b元，规定一旦A赢b元或输光了a元时，赌博就停止。下面我们讨论A最终赢的概率是多大？而A最终输的概率又是多大？

上述的输光问题可以抽象为另一类型的一维直线整数点上随机游动的模型，即为两端带有吸收壁的随机游动。假设质点每次仍以概率p和q分别向右和向左移动一位，且每次游动相互独立，但假定质点在时刻t=0时，位于$X_0=n$， $0\le n\le a+b$，而在0及$a+b$处各有一个吸收壁，即质点一旦移动到达0或$a+b$，质点永远停留在该处。我们来求质点最终在0被吸收的概率，用的是差分方程法。

若以p_n 记质点的初始位置为$X_0=n$ 而随机游动在0点被吸收的概率，以q_n记初始位置为n在$a+b$点被吸收的概率，显然 $p_{a+b}=0$， $q_{a+b}=1$ ，$p_0=1$， $q_0=0$ (1.3)

如果某时刻质点位于$X=n$ ，则它要被0吸收有两种方式来实现：一种是接下去一次是向右而最后被0吸收，另一种是接下去一次是向左的而最后被0吸收。所以按全概率公式有

$$p_n=p\,|\,p_{n+1}+q\,|\,p_{n-1} \qquad n=1，2，\equiv a+b-1 \tag{1.4}$$

这样，我们得到了关于p_n 的一个有限差分方程(1.4)，再利用边界条件(1.3)，就可以求得p_n，下面我们来求解。把(1.4)改写成

$$q(p_n-p_{n-1})=p(p_{n+1}-p_n) \qquad n=1,2,\propto, a+b-1 \tag{1.5}$$

我们先就$p=q=1/2$ 来解上面的方程，这相应于对称随机游动的场合，这时方程成为$p_n-p_{n-1}=p_{n+1}-p_n$ 。若设 $p_{n+1}-p_n=p_n-p_{n-1}=\propto=p_1-p_0=c$ ， 这里c是常数。由此可得$p_n=p_0+nc$

因为$p_0=1$，$p_{a+b}=0$， 故有 $p_0-p_n=\dfrac{n}{a+b}$。 (1.6)

因此从$X_0=a$ 出发而在0被 吸收的概率为 $p_a=1-\dfrac{a}{a+b}=\dfrac{b}{a+b}$。 (1.7)

同理可求得在对称随机游动场合，从$X_0=a$出发而在$a+b$被吸收的概率为 $q_a=\dfrac{a}{a+b}$。 (1.8)

注意到 $p_a+q_a=1$ 即质点最终一定要被吸收。

在一般场合，即$p\ne q$时，由(1.5)得

$$q^n\cdot\prod_{k=1}^{n}(p_k-p_{k-1})=p^n\cdot\prod_{k=1}^{n}(p_{k+1}-p_k)$$

因此$q^n(p_1-p_0)=p^n(p_{n+1}-p_n)$ $n=0$，1，2，$\propto$， $a+b-1$， 由(1.3)， $p_0=1$， 所以

$$p_{n+1}-p_n=(\frac{q}{p})^n(p_1-1), \quad n=0，1，2，\propto， a+b-1 \tag{1.9}$$

因为：$p_{a+b}-p_n=\displaystyle\sum_{k=n}^{a+b-1}(p_{k+1}-p_k)=\sum_{k=n}^{a+b-1}(\frac{q}{p})^n(p_1-1)=(p_1-1)\frac{(\frac{q}{p})^n-(\frac{q}{p})^{a+b}}{1-\frac{q}{p}}$

由(2.3) p_{a+b}=0，故 $$p_n=(1-p_1)\frac{(\frac{q}{p})^n-(\frac{q}{p})^{a+b}}{1-\frac{q}{p}},\tag{1.10}$$

因p_0=1，故 $$1=(1-p_1)\frac{1-(\frac{q}{p})^{a+b}}{1-\frac{q}{p}},\tag{1.11}$$

由(1.10)及(1.11)立即得到 $$p_n=\frac{(\frac{q}{p})^n-(\frac{q}{p})^{a+b}}{1-(\frac{q}{p})^{a+b}},$$

因此，质点从$X_0=a$出发，在0被吸收的概率为

$$p_a=\frac{(\frac{q}{p})^a-(\frac{q}{p})^{a+b}}{1-(\frac{q}{p})^{a+b}}=\frac{1-(\frac{p}{q})^b}{1-(\frac{p}{q})^{a+b}}。\tag{1.12}$$

用同样的方法可以求得从$X_0=a$出发，而在$a+b$被吸收的概率为

$$q_a=\frac{1-(\frac{q}{p})^a}{1-(\frac{q}{p})^{a+b}}。\tag{1.13}$$

注意到$p_a+q_a=1$，因此质点最后一定要被吸收掉。

另外，表达式(1.12)及(1.13)在$p=q=1/2$时没有意义，这时解为(1.7)及(1.8)，但是(1.7)及(1.8)也可以从(1.12)及(1.13)中令p □q 并利用洛必达法则得到。

这样，我们在这部分开头所提出的问题就得到了解决。赌博过程中A的钱增加和减少时对应于质点向右和向左随机游动，A最终赢得赌博对应于质点被吸收壁 $x=a+b$吸收，所以A 赢的概率为q_a ，而A 输的概率为p_a 。

§2 首达时间的分布及其数学期望

接下来，让我们继续讨论在无限制随机游动中经常要用到的首达时间问题。

考虑一维直线随机游动中的首次通过时间(首达时间)的分布及其数学期望。首先让我们先来看何为首达时间。

设质点在一维直线上作随机游动，不妨设$X_0=0$ 。令：$T_{0b}=min\{n,\ n\geq0，X_0=0，X_n=b\}$

其中b为整数($b>0$) 。若$\{n: n\geq 0，X_0=0，X_n=b\}=\phi$ ，取$T_{0b}=+\infty$ 。则T_{0b} 表示质点零时刻从原点出发，首次到达b的首达时间。

1. T_{01} 的分布及数学期望

（1）先看两种极端情形

当$p=1$，$q=0$时，显然$P(T_{01}=1|X_0=0)=1$ ，且$ET_{01}=1$;
当$p=0$，$q=1$时，显然对于任意的i 满足$i\in N$，有: $P(T_{01}=i|X_0=0)=0$， 于是对$\forall\ n\in N$

$P(T_{01}\leq n|X_0=0)=0$， $P(T_{01}>n|X_0=0)=1$ 又 $(T_{01}=\infty)=\bigcap_{n=1}^{\infty}(T_{01}>n)$ 且$(T_{01}>n+1)\subset(T_{01}>n)$

故: $(T_{01}=\infty)=\underset{n\to\infty}{Lim}(T_{01}>n)$ 由概率连续性，有

$$P(T_{01}=\infty\mid X_0=0)=P(\underset{n\to\infty}{Lim}(T_{01}>n)\mid X_0=0)=\underset{n\to\infty}{Lim}P(T_{01}>n\mid X_0=0)=1$$

故: $E(T_{01}|X_0=0)=+\infty$.

（2）再考虑以下一般情形

设$0<p<1$， 先求T_{01}的分布律。易知 $P(T_{01}=1|X_0=0)=p$， $P(T_{01}=2n\ |X_0=0)=0$ 以下只需讨论$P(T_{01}=2n+1|X_0=0)$ 的情形.显然$(T_{01}=2n+1)$发生的必要条件是在$2n+1$次游动中，恰有$n+1$ 次向右游动，n次向左游动，因此，可设:

$$P(T_{01}=2n+1|X_0=0)=a_np^{n+1}q^n=a_np(pq)^n \tag{2.1}$$

其中a_n表从原点出发在$2n+1$时刻首次到达1的所有可能的不同路径的数目。显然，$a_0=1$，$a_1=1$， 为求 $a_n(n\geq 2)$可以有不同的方法。这里只讨论用母函数的方法。为此，我们先利用事件的分解导出 a_n的关系式，然后导出其母函数所满足的方程。将求出的表达式再展成n的级数形式，即可求出a_n。求解过程如下。

(a) 求出a_n的关系式

令 $X'_n=X_{n+1}$ $T'_{-10}=min\{n，n>0，X'_0=-1，X'_n=0\}=min\{n，n>0，X_1=-1，X_{n+1}=0\}$，$T'_{-10}$ 表示对 $X'_n=X_{n+1}$ 而言质点从 -1 点出发，首次到达 0 点的首达时间。令

$X''_n=X'_{T'_{-10}+n}$ $T''_{01}=\min\{n,n>0,X''_0=0,X''_n=1\}$，$T''_{01}$可表示对$X''_n$而言质点从0点出发，首次到达1点的首达时间。

可以证明，T_{01},T'_{-10},T''_{01} (条件)独立，且同分布(此处留给读者思考)， 注意到，当

$n \geq 1$ 时，有如下事件分解:

$$(T_{01}=2n+1)=\bigcup_{k=0}^{n-1}\{X_0=0, X_1=-1，T'_{-10}=2k+1, T''_{01}=2n-(2k+1)\}$$

即质点必须先从0点游动一步到−1点，再经过$2k+1$步从−1点首达0点，最后再经$2n-(2k+1)$步首达1点，然后对$0 \leq k \leq n-1$ 求并。又注意到$\{X_n, n \geq 0\}$具有无后效性，即$\forall\ i，j \in N$。容易证明 $P(X_{n+1}=j|X_0, X_1, \propto, X_n=i)=P(X_{n+1}=j|X_n=i)$，我们有

$$\begin{aligned}P(T_{01}=2n+1 \mid X_0=0)&=\sum_{k=0}^{n-1}P(X_1=-1 \mid X_0=0)P(T'_{-10}=2k+1 \mid X_0=0, X_1=1)\\&\cdot P(T''_{-10}=2(n-k-1)+1 \mid X_0=0, X_1=-1, \cdots X_{2k+2}=0)\\&=\sum_{k=0}^{n-1}qP(T'_{-10}=2k+1 \mid X'_0=-1)P(T''_{01}=2(n-k-1)+1 \mid X''_0=0)\end{aligned}$$

故

$$P(T_{01}=2n+1 \mid X_0=0)=\sum_{k=0}^{n-1}qP(T_{01}=2k+1 \mid X_0=0)P(T_{01}=2(n-k-1)+1 \mid X_0=0) \quad (2.2)$$

将(2.1)代入(2.2) 。化简得

$$\begin{cases} a_0=1 \\ a_n=\sum\limits_{k=0}^{n-1}a_k a_{n-k-1}. \quad n \geq 1 \end{cases} \quad (2.3)$$

(b) 求a_n

设$\{a_n，n \geq 0\}$的生成函数(母函数)为 $A(x)=\sum\limits_{n=0}^{\infty}a_n x^n \quad 0 \leq x \leq 1$ 。对(2.3)式两边同乘以x^n ，再对$n \geq 1$求和，即可得到（详细的计算请读者自己完成）：

$$x \mid [A(x)]^2-A(x)+1=0 \quad (2.4)$$

解(2.4)式，并考虑到$A(x)$为x的单调增函数（请读者补出证明），有

$$A(x)=\frac{1-\sqrt{1-4x}}{2x} \quad (2.5)$$

把 (2.5) 式展开成幂级数，则有：$A(x)=\sum\limits_{n=0}^{\infty}\frac{1}{n+1}C_{2n}^{n}x^n$ ， 即得 $a_n=\frac{1}{n+1}C_{2n}^{n}$ ，故得到，T_{01}的分布率为

$$P(T_{01}=2n+1 \mid X_0=0)=\frac{1}{n+1}C_{2n}^{n}p(pq)^n \quad n \in N \quad (2.6)$$

由以上的结果，我们可以知道：当$p \geq q$时，$P(T_{01}<\infty|X_0=0)=1$。事实上，有

$$P(T_{01}<\infty \mid X_0=0)=P(\bigcup_{n=0}^{\infty}(T_{01}=2n+1)\mid X_0=0)=\sum_{n=0}^{\infty}P(T_{01}=2n+1\mid X_0=0)$$

$$=\sum_{n=0}^{\infty}a_n p(pq)^n=pA(pq)=p\frac{(p+q)-\sqrt{(p+q)^2-4pq}}{2pq}=p\frac{p+q-(p-q)}{2pq}=1 \tag{2.7}$$

当$p<q$时，可得 $P(T_{01}<\infty \mid X_0=0)=p\dfrac{p+q-(q-p)}{2pq}=\dfrac{p}{q}<1\Rightarrow P(T_{01}=\infty \mid X_0=0)$

$=1-\dfrac{p}{q}>0\Rightarrow E(T_{01}\mid X_0=0)=+\infty$ 。 此时T_{01}不是几乎处处有限的$r.v.$!

求出了T_{01}的分布律，即可求出其期望值。仅把结果列在下面：

$$E(T_{01}\mid X_0=0)=\begin{cases}\dfrac{1}{p-q} & p>q\\ +\infty & p\le q\end{cases} \tag{2.8}$$

详细的推导留给读者作为练习。（提示：利用期望的定义及$A(x)$的关系式）

2 $T_{0b}(b\geq 2)$的分布及期望

先来分析T_{02} ，对于T_{02}的分布，可以类似于T_{01}的分布律的方法求之。但若用如下随机变量分解的方法，则更简洁。

令 $X'_n=X_{T_{01}+n}$ $T'_{12}=\min\{n,n>0,X'_0=1,X'_n=2\}$ 容易证明T_{01} 与T'_{12}条件独立且同分布，又注意到$T_{02}=T_{01}+T'_{12}$ 故 $E(T_{02}\mid X_0=0)=2E(T_{01}\mid X_0=0)=\begin{cases}\dfrac{2}{p-q} & p>q\\ +\infty & p\le q\end{cases}$

令 $P(T_{02}=2n+2|X_0=0)=b_n(pq)^{n+1}$ $B(x)=\sum_{n=0}^{\infty}b_n x^n$ $0\le x\le 1$

易知 $B(x)=[A(x)]^2$ 再利用上面的结果，有

$$B(x)=[A(x)]^2=(\frac{1-\sqrt{1-4x}}{2x})^2=\frac{1-2x-\sqrt{1-4x}}{2x^2}$$

将$B(x)$级数展开，得 $B(x)=\sum_{n=0}^{\infty}\dfrac{1}{n+2}C_{2n+2}^{n+1}x^n$ ，故有 $b_n=\dfrac{1}{n+2}C_{2n+2}^{n+1}$，即有

$$P(T_{02}=2n+2\mid X_0=0)=\frac{1}{n+2}C_{2n+2}^{n+1}(pq)^{n+1} \tag{2.9}$$

对于$T_{0b}\,(b>2)$的情况，可仿照T_{02}的变量分解方法，类似进行处理。

以上我们着重讨论了直线上一维随机游动的两种特殊情况。还有一些较为一般的随机游动留到第九章去讨论。

§3 离散鞅简述*

在第一节的讨论中我们已经由公平赌博引出了鞅(*Martingale*)的概念，而鞅论目前已成为研究概率论理论及应用概率论和其它随机过程的有力工具。在统计、序贯决策、最优控制以及随机微分方程等方面均得到了广泛的应用。鞅论的发展与现今的竞争社会是分不开的。有奖彩票， 保险， 投资建设等均与鞅论有关。

1 定义与例子

定义 3.1 过程$\{X_n，n\geq 0\}$是鞅，如果 $\forall n\geq 0$， 有

1）$E|X_n|<\infty$，

2）$E(X_{n+1}|\ X_0，X_1，\propto，X_n)=X_n\quad a.s.$ (几乎处处)

鞅的背景来源于公平赌博。上式表明，如第 n 次赌后资金为 X_n，则第 $n+1$ 次赌博后的平均资金恰等于 X_n，即每次赌博胜负机会均等。

有时$\{X_n，n\geq 0\}$不能直接观察，而只能观察另一过程$\{Y_n，n\geq 0\}$。故我们作如下定义:

定义 3.2 设有二过程 $\{X_n，n\geq 0\}$ 及$\{Y_n，n\geq 0\}$， 称$\{X_n，n\geq 0\}$关于$\{Y_n，n\geq 0\}$是鞅，如果:

1） $E|X_n|<\infty$，

2） $E(X_{n+1}|\ Y_0，Y_1，\propto，Y_n)=X_n\quad a.s.$

注： $EX_{n+1}=E[E(X_{n+1}|\ X_0，X_1，\propto，X_n)]=EX_n=EX_0$. 这说明鞅$\{X_n，\ n\geq 0\}$在任何时刻的期望值均相等。

下面介绍一些鞅的典型例子。

例 3.1 独立同分布随机变量之和

设 $Y_0=0$，$\{Y_n, n\geq 1\}$独立同分布，$EY_n=0$，$E|Y_n|<\infty$，$X_0=0$，$X_n=\sum_{i=1}^{n}Y_i$， 则$\{X_n,n\geq 0\}$关于 $\{Y_n，n\geq 0\}$是鞅。

证明: 因为 $E|X_n|=E|\sum_{i=1}^{n}Y_i|<\infty$ 且 $E(X_{n+1}|Y_0，Y_1，\propto，Y_n)=E(X_n+Y_{n+1}|\ Y_0，Y_1，\propto，Y_n)=E(X_n|\ Y_0，Y_1，\propto，Y_n)+E(Y_{n+1}|\ Y_0，Y_1，\propto，Y_n)=X_n$。

例 3.2 和的方差

设 $Y_0=0$， $\{Y_n，n\geq 1\}$独立同分布，$EY_n=0$，$EY_n^2=\sigma^2$，$X_0=0$， $X_n=(\sum_{i=1}^{n}Y_i)^2-n\sigma^2$，则$\{X_n，n\geq 0\}$关于$\{Y_n，n\geq 0\}$是鞅。

证明： 因为

$$E|X_n| = E|(\sum_{i=1}^{n} Y_i)^2 - n\sigma^2| \le E|(\sum_{i=1}^{n} Y_i)^2| + n\sigma^2$$

$$= E(\sum_{i=1}^{n} Y_i^2 + \sum_{i \ne j} Y_i Y_j) + n\sigma^2 = 2n\sigma^2 < \infty$$

且

$$E(X_{n+1} | Y_0, \cdots Y_n) = E[(Y_{n+1} + \sum_{i=1}^{n} Y_i)^2 - (n+1)\ \sigma^2 | Y_0, \cdots Y_n)$$

$$= E[Y_{n+1}^2 + 2Y_{n+1}\sum_{i=1}^{n} Y_i + (\sum_{i=1}^{n} Y_i)^2 - (n+1)\sigma^2 | Y_0, \cdots Y_n]$$

$$= E[Y_{n+1}^2 | Y_0, \cdots Y_n] + 2E[Y_{n+1}\sum_{i=1}^{n} Y_i | Y_0, \cdots Y_n] + E(X_2 | Y_0, \cdots Y_n) - \sigma^2$$

$$= EY_{n+1}^2 + 2E[Y_{n+1} | Y_0, \cdots Y_n](\sum_{i=1}^{n} Y_i) + X_n - \sigma^2 = X_n$$

由以上两例知，由独立同分布随机变量的和或者和的方差所构成的序列都可以构造鞅。

例 3.3 似然比构成的鞅

设 Y_0，Y_1，∝，Y_n,∝是独立同分布随机变量序列，f_0 和 f_1 是 $p.d.f$ ，令

$$X_n = \frac{f_1(Y_0) \cdot f_1(Y_1) \cdots f_1(Y_n)}{f_0(Y_0) \cdot f_0(Y_1) \cdots f_0(Y_n)} \quad n \ge 0$$

假设$\forall y$，$f_0(y) > 0$，当 Y_n 的 $p.d.f.$为 f_0 时，则$\{X_n，n\ge0\}$关于$\{Y_n，n\ge0\}$是鞅。

证明: 因为

$$E|X_n| = E|\frac{f_1(Y_0) \cdot f_1(Y_1) \cdots f_1(Y_n)}{f_0(Y_0) \cdot f_0(Y_1) \cdots f_0(Y_n)}| < \infty\text{，且}$$

$$E(X_{n+1} | Y_0, \cdots, Y_n) = E[X_n \frac{f_1(Y_{n+1})}{f_0(Y_{n+1})} | Y_0, \cdots, Y_n] = X_n E[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}]$$

由于 $E[\frac{f_1(Y_{n+1})}{f_0(Y_{n+1})}] = \int \frac{f_1(y)}{f_0(y)} \cdot f_0(y)dy = \int f_1(y)dy = 1$，因此$\{X_n，n\ge0\}$关于$\{Y_n，n\ge0\}$是鞅。

例 3.4 *Doob* 鞅过程

设 Y_0，Y_1，∝，Y_n,∝是一随机序列. 有一随机变量 X，$E|X|<\infty$，令

$$X_n = E(X| Y_0, \propto, Y_n)$$

则$\{X_n，n\ge0\}$关于$\{Y_n，n\ge0\}$是鞅，并称之为 *Doob* 过程。

证明: 因为 $E|X_n| = E\{|E(X| Y_0, \propto, Y_n)|\} \le E\{E(|X|| Y_0, \propto, Y_n)\} = E|X| < \infty$，

$E(X_{n+1}| Y_0, \propto, Y_n) = E\ \{E(X| Y_0, \propto, Y_{n+1})| Y_0, \propto, Y_n\} = E(X| Y_0, \propto, Y_n)=X_n$.

这个例子很“奇特”， 以一系列任意随机变量为条件的条件数学期望构成鞅。

以上列举了四个鞅的例子， 许多时候用鞅可以解决原来不易解决的问题. 但关键是如何构造出鞅来。

由不公平赌博还可引出上鞅和下鞅的概念，如下：

定义 3.3 设有二过程 $\{X_n，n≥0\}$ 及$\{Y_n，n≥0\}$， 称$\{X_n，n≥0\}$关于$\{Y_n，n≥0\}$是上鞅， 如果:

1） $E|X_n|<\infty$，

2） $E(X_{n+1}| Y_0，Y_1，\propto，Y_n) \le X_n \quad a.s$，

3） X_n 是 $Y_0，Y_1，\propto，Y_n$ 的函数。

定义 3.4 设有二过程 $\{X_n，n≥0\}$ 及$\{Y_n，n≥0\}$， 称$\{X_n，n≥0\}$关于$\{Y_n，n≥0\}$是下鞅， 如果:

1） $E|X_n|<\infty$，

2） $E(X_{n+1}| Y_0，Y_1，\propto，Y_n) \ge X_n \quad a.s$，

3） X_n 是 $Y_0，Y_1，\propto，Y_n$ 的函数。

2 停时与停时定理

我们知道若$\{X_n，n≥0\}$是鞅，则：$\forall n\in N$ ，有 $EX_n=EX_0$，那么若 T 是一个取值非负整数的随机变量，在 $EX_n=EX_0$ 中以 T 代替 n，请读者从直观上猜想，是否有 $EX_T=EX_0$？例如，本章第一节中对称随机游动。由(1.2)式知，当 $p=q$ 时，$\{X_n，n≥0\}$是鞅，有 $EX_n=EX_0$。然而当 T 为首达时间 T_{01} 时，由(2.7)式有 $P(T_{01}<\infty \mid X_0=0)=1$。再由 $X_{T_{01}}$ 定义，可得 $P(X_{T_{01}}=1)=1$，故 $E\,X_{T_{01}}$ =1。而 EX_0＝0，所以 $E\,X_{T_{01}} \ne EX_0$。那么什么时候二者相等呢？为此先引出停时的概念。停时是一个不依赖于“将来”的随机时间，先给出粗略的直观定义。

定义 3.5 设取值为非负整数(包括+∞)的随机变量 T，及随机序列$\{Y_n，\ n≥0\}$. 若对 $n≥0$， 事件$\{T=n\}$的示性函数 $I_{\{T=n\}}$仅是 $Y_0，Y_1，\propto，Y_n$ 的函数， 则称 T 是关于$\{Y_n，n≥0\}$的停时(*stopping time*) (或称马氏时间)。

例：上段对称随机游动中的 T_{01} 关于$\{X_n，n≥0\}$是停时，因 $\forall n\in N$ ，有 $I_{(T_{01}=n)}=$

$I_{(X_k\ne 1, 1\le k\le n-1, X_n=1)}$，显然它是 $(X_k, 1\le k\le n)$ 的函数，故 T_{01} 关于$\{X_n，n≥0\}$是停时。

例 3.5 $\{N(t),\ t\geq 0\}$是参数为λ的时齐 *Poisson* 过程，$S_0=0$，S_n为第 n 个事件发生时刻，则 $N(t)$关于$\{S_n,\ n\geq 0\}$不是停时，但 $N(t)+1$ 关于$\{S_n,\ n\geq 0\}$是停时。

证明：请读者自己完成。

停时有以下基本特性：

1）$T=k$ (k 是一常数) 是一个停时；

2）设 T，σ是关于$\{Y_n,\ n\geq 0\}$的两个停时，则 $T+\sigma$，$T\vee\sigma=max(T,\sigma)$，$T\wedge\sigma=min(T,\sigma)$均是停时。

下面介绍有关停时定理(*Optional stopping theorem* 或 *Optional Sampling theorem*).及其应用。我们将不加证明的给出一个较为一般的停时定理及其推论：

定理 3.1 (停时定理) 设$\{X_n,\ n\geq 0\}$是鞅，T 是停时，若

1) $P(T<\infty)=1$;

2) $E|X_T|<\infty$;

3) $\lim\limits_{n\to\infty} E|X_n\cdot I_{\{T>n\}}|=0$.

则

$$EX_T=EX_0.$$

推论 设 $Y_0=0$，$\{Y_k,\ k\geq 1\}$独立同分布，$EY_k=\mu$，$DY_k=\sigma^2<\infty$，$S_0=0$，$S_n=\sum_{k=1}^{n} Y_k$，$X_n=S_n-n\mu$. 若 T 为停时，$ET<\infty$，则 $E|X_T|<\infty$，且 $EX_T=ES_T-ET=0$。

如何在实际中应用停时定理与推论呢？这里介绍一个应用停时定理的例子。

例 3.6 随机游动

$\{Y_n,\ n\geq 0\}$，$\{X_n,\ n\geq 0\}$，T_{0j}的定义均与第一节相同，以下令

$T=min\{n{:}X_n=a$ 或 $X_n=b\}$，$b>0$ 为正整数，$a<0$ 为负整数，

$V_a=P(X_T=a\mid X_0=0)$，V_a表从 0 出发先到达 a 的概率。

则 $P(X_T=b\mid X_0=0)=1-V_a$。

若以赌博(或投资)为背景。设甲乙两人赌博，$|a|$表甲原有的资金，Y_n 表甲第 n 次得到的钱，b 表乙原有的资金。那么 V_a表示甲先输光的概率。

分二种情况讨论

1) 当 $p=q=\dfrac{1}{2}$ 时

由(2.7)式的计算知：$P(T_{01}<\infty\mid X_0=0)=1$。由 $X_{T_{01}}$ 定义知，$X_{T_{01}}=1$ 故 $EX_{T_{01}}=1$，而 $EX_0=0$，所以 $EX_{T_{01}}\neq EX_0$。易知$\{X_n,\ n\geq 0\}$关于$\{Y_n,\ n\geq 0\}$是鞅，T_{01} 关于$\{Y_n,\ n\geq 0\}$是停时. 由定理 3.1 的推论知，$ET_{01}<\infty$ 不成立 (因为若 $ET_{01}<\infty$，则 $EX_{T_{01}}=EX_0=0$)，

故 $ET_{01}=+\infty$。

可以验证 T 满足定理 3.1 的条件，故有：$EX_T=EX_0=0$。但 $EX_T=V_a a+(1-V_a)b=0$，解之，得：$V_a=\dfrac{b}{|a|+b}$，这就是甲先输光的概率。同理：$V_b=1-V_a=\dfrac{|a|}{|a|+b}$。

如何求 $E(T|X_0=0)$，即任何一方输光的平均时间是多少呢？这需要构造一个合适的鞅来解决。

设 $Z_n=X_n^2-n$，不难验证它关于$\{Y_n,\ n\geq 0\}$是鞅，详细的推导留给读者。

由于 T 为停时，且 $ET<\infty$。另外，可以验证：Z_n 与 T 满足定理 3.1 的条件，故有：$EZ_T=EZ_0=0$.

而由于 $EZ_T=E(X_T^2-T)=EX_T^2-ET$，因此：

$$ET=EX_T^2=a^2V_a+b^2(1-V_a)=a^2\cdot\frac{b}{|a|+b}+b^2\cdot\frac{|a|}{|a|+b}=|a|b$$

2）当 $p-q=\mu>0$ 时，这是不公平赌博，每次甲方赢的概率大，这时的 V_a 是多少呢？为此仍需要构造鞅。注意到此时 $EY_n=p-q=\mu$. 令 $U_n=X_n-n$。容易验证：$\{U_n,\ n\geq 0\}$关于$\{Y_n,\ n\geq 0\}$是鞅。另外，可以验证：U_n 与 T 满足定理 3.1 的条件，有：$EU_T=EU_0$，故 $EX_T=\mu ET$。对于 T_{01}，我们知道它也是停时。同样可以验证：U_n 与 T_{01} 满足定理 3.1 的条件，故：$EX_{T_{01}}=\mu ET_{01}$，即：

$$ET_{01}=\frac{EX_{T_{01}}}{\mu}=\frac{1}{p-q}$$

在此我们得到了第一节中已有的结论，但此处的计算要简洁的多，这就是停时定理的威力。

令 $V_n=(\frac{q}{p})^{X_n}$，由于 $p>q$，故 $0<\frac{q}{p}<1$. 下面验证$\{V_n,\ n\geq 0\}$关于$\{Y_n,\ n\geq 0\}$是鞅. 因为

$$E|V_n|=E|(\frac{q}{p})^{X_n}|=E(\frac{q}{p})^{X_n}=\prod_{k=1}^{n}E(\frac{q}{p})^{Y_k}=1<\infty$$

$$E(V_{n+1}|Y_0,\cdots Y_n)=E[(\frac{q}{p})^{X_{n=1}}|Y_0,\cdots Y_n]=E[(\frac{q}{p})^{X_n}\cdot(\frac{q}{p})^{Y_{n+1}}|Y_0,\cdots Y_n]$$

$$=(\frac{q}{p})^{X_n}\cdot E[(\frac{q}{p})^{Y_{n+1}}]=(\frac{q}{p})^{X_n}\cdot[\frac{q}{p}\cdot p+(\frac{q}{p})^{-1}\cdot q]=(\frac{q}{p})^{X_n}=V_n$$

可以验证：V_n 与 T 满足定理 3.1 的条件，故 $EV_T=EV_0$，另外，由于 $EV_0=1$，因此 $EV_T=1$。

仍记 $V_a=P(X_T=a \mid X_0=0)$，有 $EV_T=V_a(\frac{q}{p})^a+(1-V_a)(\frac{q}{p})^b=1$ 解之，得

$$V_a=\frac{1-(\frac{q}{p})^b}{(\frac{q}{p})^a-(\frac{q}{p})^b}<\frac{b}{|a|+b}$$

故从 0 状态出发，当 $p>q$ 时首达 a 的概率小于 $p=q$ 时首达 a 的概率。

在以上论证过程中，满足停时定理条件的证明均省略了，这主要是因为其中有些已超出了本书的范围。有兴趣的读者可以参看[11]。

练习题

7.1 设$\{Y_n, n\ge 1\}$ *i.i.d.* $P(Y_n=1)=p\ge 0, P(Y_n=-1)=q\ge 0, p+q=1,$

$p-q>0, Y_0=0,$ $X_0=0, X_n=\sum_{k=1}^n Y_k, U_n=X_n-n(p-q),$ $V_n=(q/p)^{X_n},$

$W_n=U_n^2-n[1-(p-q)^2]$。(1)证$\{U_n,n\ge 0\},\{V_n,n\ge 0\},\{W_n,n\ge 0\}$关于$\{Y_n,n\ge 1\}$是鞅；(2)证$\{X_n,n\ge 0\}$关于$\{Y_n,n\ge 1\}$是下鞅；(3)求$U_m$与$U_n$的相关系数；(4)求$E(U_2\mid X_2)$的分布律并证明$E(U_{n+k}\mid X_n)=U_n,\forall x\ge 0$；（5）求$E(V_8\mid X_7=3)$。

7.2 设$Y_0=0,\{Y_n,n\ge 1\}$ *i.i.d.*，（1）若$EY_n=0, EY_n^2=\sigma^2, X_0=0,$ $X_n=(\sum_{k=1}^n Y_k)^2-n\sigma^2$，证$\{X_n,n\ge 0\}$关于$\{Y_n,n\ge 1\}$是鞅；(2)若$Y_n\sim N(0,\sigma^2), X_n=\exp\{\frac{\mu}{\sigma^2}\sum_{k=1}^n Y_k-\frac{n\mu^2}{2\sigma^2}\}$，$n\ge 1, X_0=0$。证明$\{X_n,n\ge 0\}$关于$\{Y_n,n\ge 1\}$是鞅。

7.3 设随机变量序列$\{X_n,n\ge 0\}$，满足$E|X_n|<\infty$，且

$E[X_{n+1}\mid X_0,X_1,\cdots,X_n]=\alpha X_n+\beta X_{n-1}, n>0, \alpha,\beta>0, \alpha+\beta=1$。

令$Y_n=aX_n+X_{n-1}, n\ge 1, Y_0=X_0$。试选择合适的$a$使得$\{Y_n,n\ge 1\}$关于$\{X_n,n\ge 0\}$是鞅。

7.4 设$\{X_n,n\ge 0\}$关于$\{Y_n,n\ge 1\}$是鞅。证明对任何整数集$k\le l<m, X_m-X_l$与X_k不相关，即$E[(X_m-X_l)X_k]=0$。

7.5 设$\{X_n,n\ge 0\}$是鞅，而$\{\xi_i,i\ge 0\}$由下式部分和确定，$X_n=\sum_{i=0}^n \xi_i$。试证：

$\forall j\ne i,\ E(\xi_i\xi_j)=0$。

7.6 设Y_0服从 (0, 1) 上均匀分布；且给定Y_n时，Y_{n+1}是$(1-Y_n,1]$上的均匀分布。令$X_0=Y_0, X_n=2^n\prod_{k=1}^{n}\left[(1-Y_k)/Y_{k-1}\right], n=1,2,\cdots$。证明$\{X_n, n\geq 0\}$关于$\{Y_n, n\geq 1\}$是鞅。

7.7 设$\{X_n, n\geq 0\}$是 M.C.状态空间$S=\{0,1,2,\cdots,N\}$。

(1) 若$p_{ij}=C_N^j\left(i/N\right)^j\left(1-i/N\right)^{N-j}, V_n=X_n\left(N-X_n\right)/\left(1-N^{-1}\right)^n$。试证:

$\{X_n, n\geq 0\}$ 与$\{V_n, n\geq 0\}$关于$\{X_n, n\geq 0\}$是鞅。

(2) 若$p_{ij}=C_{2i}^jC_{2N-2i}^{N-j}/C_{2N}^N, W_n=X_n(N-X_n)/\lambda^n, n\geq 0$。试确定$\lambda$，使

$\{W_n, n\geq 0\}$关于$\{X_n, n\geq 0\}$是鞅。

7.8 设$\{X_n, n\geq 0\}$是 M.C.，$S=\{0,1,2,\cdots,N\}$，$P=(p_{ij}), P^{(m)}=(p_{ij}^{(m)})$。且$u(i,n)$对$\forall i,n\in S$,，满足$u(i,n)=\sum_{k=0}^{\infty}u(k,n+m)p_{ik}^{(m)}$。记$U_n=u(X_n,n)$， 证明：$\{U_n, n\geq 0\}$关于$\{X_n, n\geq 0\}$是鞅。

7.9 设$\{X_n, n\geq 0\}$关于$\{Y_n, n\geq 1\}$是鞅，c 是任意常数。（1）若$E\left|X_n\vee c\right|<+\infty$，则$\{X_n\vee c, n\geq 0\}$是下鞅；（2）若$EX_n^+<+\infty$, 则$\{X_n^+, n\geq 0\}$是下鞅。

7.10 设$\{X_n, n\geq 0\}$关于$\{Y_n, n\geq 1\}$是上鞅。（1）若$E\left|X_n\wedge c\right|<+\infty$,则

$\{X_n\wedge c, n\geq 0\}$是上鞅；（2）若$EX_n^->-\infty$, 则$\{X_n^-, n\geq 0\}$是上鞅。

7.11 设$Z\sim U[0,1], Y_n=\sum_{k=0}^{2^n-1}k/2^n\, I_{(k/2^n\leq Z<(k+1)/2^n)}, n\geq 1$，求：

$EY_1, E(Y_1Y_2), E(Y_1\mid Y_2)$，$E(Y_n\mid Y_1,\cdots,Y_{n-1})$。

7.12 Y_n, X_n 同 7.1 题，$T_{01}=\min\{n: n>0, X_n=1\}$，$X_n'=X_{T_{01}}+n$，$T_{12}'=\min\{n: n>0, X_n'=2\}$，试证：$T_{01}$与$T_{12}'$条件独立且同分布。

注：其中第7.7、7.8两题可等学完第九章后再做。

第八章 *Brown* 运动

Brown 运动最初是由英国生物学家 *R.Brown* 于 1827 年根据花粉微粒在液面上做"无规则运动"的物理现象提出的。1918 年 *Wiener* 对这一现象在理论上作了精确的数学描述，并进一步研究了 *Brown* 运动轨道的性质，提出了 *Brown* 运动空间上的测度与积分，使得对 *Brown* 运动及其泛函的研究得到迅速而深入的发展。*Brown* 运动作为具有连续时间参数和连续空间参数的一个随机过程，是一个最基本、最简单同时又是最重要的过程。许多其它的过程常常可以看作是它的泛函或某种意义上的推广。今天，*Brown* 运动及其推广已广泛地出现在许多纯科学与应用科学领域。同时，由于 *Brown* 运动与微分方程（如热传导方程）有密切的联系，它又成为概率与分析联系的重要渠道。在这一章里，仅对 *Brown* 运动作一简要的介绍 。

§1 定义和性质

1. *Brown* 运动的定义

我们从一维直线上的随机游动引出布朗运动。

i) 先离散化。 让我们先考虑在一直线上的简单、对称的随机游动。设质点每隔 Δt 时间，随机地以概率 $p=\frac{1}{2}$ 向右移动 $\Delta x>0$ 单位，或以概率 $q=1-p=\frac{1}{2}$ 向左移动 Δx 单位，且每次移动相互独立，记

$$X_i=\begin{cases}1, & \text{第}i\text{次质点向右移动}\\ -1, & \text{第}i\text{次质点向左移动}\end{cases}$$，$X(t)$=t 时刻质点的位置。

设 t=0 时，$X(0)$=0，那么有 $X(t)=\Delta x(X_1+X_2+\cdots+X_{[\frac{t}{\Delta t}]})$，其中 $[\frac{t}{\Delta t}]$ 表示不超过 $\frac{t}{\Delta t}$ 的最大整数。显然 $X(t)=\Delta x(\sum_{k=1}^{[\frac{t}{\Delta t}]}X_k)$ 是独立同分布的 *r.v.*之和，因而它们在互不重叠的时间区间上，其增量相互独立，且 $EX_i=0$，$DX_i=EX_i^2=1$。故此时 $EX(t)=0$，$DX(t)=(\Delta x)^2\cdot[\frac{t}{\Delta t}]$。

以上简单随机游动可作为微小粒子在直线上作不规则运动的近似。而实际粒子的不

规则运动是连续进行的，为此，考虑 $\Delta t\to 0$ 时极限情形。

ii) 精确化。 通常情况下，令 $\Delta t\to 0, \Delta x\to 0$。由实验得知，在许多情形下，有 $\Delta x = C\sqrt{\Delta t}$ ($C>0$ 为常数)，下面假定在 $\Delta x = C\sqrt{\Delta t}$ 的条件下，推出其极限情形。

显然，当 $\Delta t\to 0$ 时，$EX(t)=0$；而 $\underset{\Delta t\leftarrow 0}{Lim}\, DX(t) = \underset{\Delta t\leftarrow 0}{Lim}(\Delta x)^2[\frac{t}{\Delta t}] = C^2 t$

另一方面， $X(t) = \Delta x(X_1 + X_2 + \cdots + X_{[\frac{t}{\Delta t}]})$ 可看作是独立同分布的随机变量之和，即 $X(t)$看作是许多微小的随机变量 $C\sqrt{\Delta t}X_1, \cdots C\sqrt{\Delta t}X_{[\frac{t}{\Delta t}]}$ 之和；故当 $\Delta t\to 0$ 时，由中心极限定理知，$X(t)$趋向于正态分布，即 $\Delta t\to 0$ 时，$X(t)\sim N(0,C^2t)$。若$\forall x\in R$，$t>0$， $\Phi(x)$为标准正态函数，有

$$\underset{\Delta t\to 0}{Lim}\, P\left\{\frac{\sum_{i=0}^{[\frac{t}{\Delta t}]}\Delta x\cdot X_i - 0}{\sqrt{C^2 t}} \le x\right\} = \Phi(x)$$

等价于： $\underset{\Delta t\to 0}{Lim}\, P\left\{\frac{X(t)}{\sqrt{C^2 t}} \le x\right\} = \Phi(x) = \frac{1}{2\pi}\cdot\int_{-\infty}^{x} e^{-\frac{u^2}{2}}du$ 。

有了上述简单随机游动的极限描述，可以引出下面的

定义 1.1 若一族随机变量 $\{X(t), t\ge 0\}$ 满足

i) $\{X(t), t\ge 0\}$ 在不重叠的区间上，其增量相互独立；

ii) $\forall s,t>0$, $X(s+t)-X(s)\sim N(0,C^2t)$，即 $X(s+t)-X(s)$是期望为 0，方差为 C^2t 的正态分布；

iii) $X(t)$关于 t 是连续函数。

则称$\{X(t), t\ge 0\}$ 是 *Brown* 运动(*B.M.* 或 *Wiener Process*)。

以后如无特别声明，约定 $X(t)=0$。

当 $C=1$ 时，称$\{X(t), t\ge 0\}$为标准 *Brown* 运动，记为$\{B(t), t\ge 0\}$。它在 t 点的 *p.d.f.*记为

$$p(x,t) = \frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}} \tag{1.1}$$

根据上面 *Brown* 运动的特性可知 $B(t_2)-B(t_1)\sim N(0,t_2-t_1)$ ，故其 *p.d.f.* 为

$$p(x,t_2-t_1) = \frac{1}{\sqrt{2\pi(t_2-t_1)}}\exp[-\frac{x^2}{2(t_2-t_1)}] \tag{1.2}$$

2. *Brown* 运动的性质

定义 1.2 随机过程$\{X(t),t\in T\}$，若对任意的 $t_i\in T,i=1,2,\cdots n$，$(X(t_1),X(t_2)\cdots X(t_n))$的联合分布为 n 维正态分布，则称$\{X(t),t\in T\}$为正态过程。

定理 1.1 *Brown* 运动是正态过程。设$\{B(t),\ t\geq 0\}$为标准 *B.M.*，则当 $t_0=0,B(0)=0$ 时 $\forall 0=t_0<t_1<\cdots<t_n,(B(t_1),B(t_2),\cdots,B(t_n))$ 的 *j.p.d.f.*为

$$g(x_1,x_2,\cdots x_n;t_1,t_2,\cdots t_n)=\prod_{i=1}^{n}p(x_i-x_{i-1},t_i-t_{i-1}) \tag{1.3}$$

其中: $p(x,t)=\dfrac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}$ 且 $x_0=0$ 。

证明：利用 *B.M.*的增量独立性。

令 $Y_1=B(t_1),\ Y_i=B(t_i)-B(t_{i-1}),2\leq i\leq n$，则 $B(t_i)=\sum_{k=1}^{i}Y_k$；由 *B.M.*的定义知 $Y_1,Y_2,\cdots,Y_n$ 相互独立,$Y_n\sim N(0,t_i-t_{I-1})$，则其联合密度函数为

$$f(y_1,y_2,\cdots y_n)=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi(t_i-t_{i-1})}}\exp[-\frac{x^2}{2(t_i-t_{i-1})}]$$

由于 $B(t_i)=\sum_{k=1}^{i}Y_k$ ，故由第三章命题 6.1 得 $(B(t_1),B(t_2),\cdots,B(t_n))$的 *j.p.d.f.*为

$$g(x_1,x_2,\cdots x_n;t_1,t_2,\cdots t_n)=f(y_1,y_2,\cdots y_n)\cdot|J|$$

且易得 $|J|=1$，故

$$\begin{aligned}g(x_1,x_2,\cdots x_n;t_1,t_2,\cdots t_n)&=\prod_{i=1}^{n}\frac{1}{\sqrt{2\pi(t_i-t_{i-1})}}\exp[-\frac{(x_i-x_{i-1})^2}{2(t_i-t_{i-1})}]\\&=\prod_{i=1-}^{n}p(x_i-x_{i-1},t_i-t_{i-1}).\end{aligned}$$

定理得证。

在得出了 $(B(t_1), B(t_2),\cdots, B(t_n))$ 的 *j.p.d.f.*之后，可得以下推论。

推论 1 $\forall 0<t_1<t_2<\cdots<t_{n-1}<t_n$,在 $B(t_1)=x_1,B(t_2)=x_2,\cdots, B(t_{n-1})=x_{n-1}$ 下 $B(t_n)$的 *c.p.d.f.*为

$$f(x_n\,|\,x_1,x_2,\cdots x_{n-1})=\frac{1}{\sqrt{2\pi(t_n-t_{n-1})}}\exp[-\frac{(x_n-x_{n-1})^2}{2(t_n-t_{n-1})}] \quad . \tag{1.4}$$

推论 2

$$E(B(t_n)|\ B(t_1)=x_1,\ B(t_2)=x_2,\cdots,\ B(t_{n-1})=x_{n-1}\ )=x_{n-1} \tag{1.5}$$

$$D(B(t_n)|\ B(t_1)=x_1,\ B(t_2)=x_2,\cdots,\ B(t_{n-1})=x_{n-1})=t_n-t_{n-1} \tag{1.6}$$

推论 1 和 2 可以通过求 *c. p.d.f.* 的方法很容易求得，这里不再赘述。

推论 1 表明 *Brown* 运动具有这样一条性质：即已知现在，将来与过去条件独立。但是, 对于 $t_1<t_2<t_3$, 如果在给定 $B(t_1)$与 $B(t_3)$下求 $B(t_2)$的 *c.p.d.f.*，结果会怎样呢？

先看看取 $t_1=0, t_3=1, t_2=t, 0<t<1$ 的情形。由式（1.3）知，在 $B(0)=0$ 下, $B(t), B(1)$的 *j.p.d.f.* 为

$$f(x,y)=\frac{1}{2\pi\sqrt{t(1-t)}}\exp\{-\frac{1}{2}[\frac{x^2}{t}+\frac{(y-x)^2}{1-t}]\}。$$

又在 $B(0)=0$ 下, $B(1)$的 *p.d.f.*为 $g(y)=\frac{1}{\sqrt{2\pi}}\exp(-\frac{y^2}{2})$，从而在 $B(0)=B(1)=0$ 下，$B(t)$的条件概率密度为： $f_{B(t)|B(0)=B(1)=0}(x)=\frac{f(x,y)}{g(y)}=\frac{1}{\sqrt{2\pi t(1-t)}}\exp[-\frac{1}{2}\frac{x^2}{t(1-t)}]$，仍为正态分布。可以看出

$$E[B(t)|B(0)=B(1)=0]=0,\quad E[B^2(t)|B(0)=B(1)=0]=t(1-t)。$$

对于一般的情形, 我们有以下

定理 1.2　对 $t_1<t<t_2$，给定 $B(t_1)=a, B(t_2)=b, B(0)=0$, 则 $B(t)$的 *c.p.d.f.*是一正态密度，其均值为：$a+(b-a)(t-t_1)(t_2-t_1)^{-1}$，方差为：$(t_2-t)(t-t_1)(t_2-t_1)^{-1}$。

证明：$E[B(t)|B(t_1)=a, B(t_2)=b]=E[B(t)-B(t_1)+B(t_1)|B(t_1)-B(0)=a, B(t_2)-B(t_1)=b-a]$

$$=E[B(t)-B(t_1)|B(t_2)-B(t_1)=(b-a)]+a\overset{*}{=}\frac{t-t_1}{\sqrt{(t-t_1)(t_2-t_1)}}\cdot\sqrt{\frac{t-t_1}{t_2-t_1}}[(b-a)-0]+a$$

$$=a+(b-a)(t-t_1)(t_2-t_1)^{-1}$$

上式*处用到了第四章例 5.5 的(5.8)式（二元正态条件期望的公式）。条件方差的证明与上例类似，留作练习。提示：先求相应的 *c.p.d.f.*。

定理 1.3　设 $\{B(t),t\geq 0\}$是正态过程，轨道连续，$B(0)=0$，$\forall s,t>0$，有 $EB(t)=0$, $E[B(s)B(t)]=t\wedge s$，则$\{B(t),t\geq 0\}$是 *B.M.*。反之亦然。

证明：先证充分性。

由定理 1.1 知, 若$\{B(t),t\geq 0\}$是 *BM*, 则它是正态过程。 同时由 *Brown* 运动定义知, $EB(t)=0$, 且轨道连续。设 $0<s\leq t$ 则

$E[B(s)B(t)]=E[(B(t)-B(s)+B(s))B(s)]=E[(B(t)-B(s))B(s)]+E[B^2(s)]$

$=E[B(t)-B(s)]E[B(s)]+s=s$。所以

$$E[B(s)B(t)]= t\wedge s，E[B(t)]=0 \qquad (*)$$

充分性得证。

再证必要性。

当$\{B(t),t\geq 0\}$是正态过程, 且满足（*）式时，那么$\forall\ s,t>0$，有

$$E[B(t)-B(s)]=E(B(t))-E(B(s))=0，$$

$$E[B(t)-B(s)]^2=E[B^2(t)]+E[B^2(s)]-2E[B(t)B(s)]=t+s-2(t\wedge s)=|t-s|。$$

而$\forall s_1<t_1\leq s_2<t_2$, 有

$E[\{B(t_1)-B(s_1)\}\{B(t_2)-B(s_2)\}]=E[B(t_1)B(t_2)]-E[B(t_1)B(s_2)]-E[B(s_1)B(t_2)]+E[B(s_1)B(s_2)]$

$=t_1-t_1-s_1+s_1=0$。

再由正态分布知：不相关即相互独立, 所以 $B(t)$是独立增量过程。且 $B(t)-B(s)\sim N(0,|t-s|)$，又$\{B(t),t\geq 0\}$轨道是连续的，得$\{B(t),t\geq 0\}$是 *B.M.*。必要性得证。

这样就得到了判断一个正态随机过程是否为 *Brown* 运动的充分必要条件, 从而得出一系列很有用的结论。

定理 1.4 设$\{B(t),t\geq 0\}$是 *Brown* 运动，则

1）$\{B(t+\tau)-B(\tau),t\geq 0\}$，$\forall\tau\geq 0$；

2）$\{\frac{1}{\sqrt{\lambda}}B(\lambda t),t\geq 0\},\lambda>0$；

3）$\{tB(\frac{1}{t}),t\geq 0\}$,其中$tB(\frac{1}{t})|_{t=0}\triangleq 0$；

4）$\{B_{T-S}-B_T,0\leq S\leq T\},T>0$。

仍为 *Brown* 运动。

证明：利用定理 1.3 的结论易证，详细的推导留给读者。

至此, 我们已经研究了不少有关 *Brown* 运动的重要性质和定理。那么, 它和我们上一章所讨论的鞅有什么关系呢？这是一个很有价值的问题。我们先引入连续参数鞅的概念。

定义 1.3 随机过程$\{X(t),t\geq 0\}$, 若$\forall\ 0\leq t_0<t_1<\cdots<t_n<t$, 有 $E|X(t)|<\infty$, 且：

$$E(X(t)|X(t_0), X(t_1),\cdots, X(t_n))=X(t_n) \quad (a.s.)$$

则称$\{X(t),t\geq 0\}$为鞅。

定理 1.5 设$\{B(t),t\geq 0\}$为 *Brown* 运动, 则$\{B(t),t\geq 0\}$, $\{B^2(t)-t, t\geq 0\}$都是鞅。

证明：利用 *Brown* 运动的增量独立性不难证明，详细的推导留给读者。

由以上结论可知, *Brown* 运动本身既是马氏正态过程, 又是连续鞅。这个结果很别致, 但并不奇怪。因为我们分别讨论的 *Poisson* 过程、 *Markov* 链、鞅、*Brown* 运动等随机过程, 不过是对一些随机过程的某些方面的特殊性质进行了专门的、分类的讨论, 而并不排斥这些性质可以交叉、共存于一个随机过程中。在介绍这些概念时只能“串行”进行, 但实际上要能够“并行”应用, 融会贯通。

§2 首中时(*Hitting Time*)与最大值(*Maximarn Value*)的分布

设$\{B(t),t\geq 0\}$为标准 *Brown* 运动。不妨设 $B(0)=0$，记 $T_a=min\{t: t>0,B(t)=a\}$，T_a 表示首次击中 a 的时间(首中时)；要研究 $P(T_a\leq t)$有多大，仍然从事件的等价性入手。

对$\forall t>0$，记 $M_t=\max\limits_{0\leq u\leq t}B(u)$ 表示$[0,t]$上的最大值。显然存在下述事件等价关系

$$\{T_a\leq t\}=\{M_t\geq a\} \qquad (2.1)$$

则有 $P\{T_a\leq t\}=P\{M_t\geq a\}$。 为求 $P\{T_a\leq t\}$, 注意到

$$P(B(t)\geq a)=P(B(t)\geq a|T_a\leq t)P(T_a\leq t)+P(B(t)\geq a|T_a>t)P(T_a>t)。$$

显然，$P(B(t)\geq a|T_a>t)=0$，又由 *Brown* 运动的对称性知，在 $T_a\leq t$ 条件下，即 $B(T_a)=a$ 时，$(B(t)\geq a)$与$(B(t)<a)$ 是等可能的，即

$$P(B(t)\geq a|T_a\leq t)=P(B(t)<a|T_a\leq t)=\frac{1}{2}$$

故 $P(T_a\leq t)=2P(B(t)\geq a)$。

于是当 $a>0$ 时，$P(T_a\leq t)=\dfrac{2}{\sqrt{2\pi t}}\displaystyle\int_a^{+\infty}e^{-\frac{u^2}{2t}}du=\sqrt{\frac{2}{\pi}}\int_{\frac{a}{\sqrt{t}}}^{+\infty}e^{-\frac{x^2}{2}}dx=2(1-\Phi(\frac{a}{\sqrt{t}}))$

当 $a<0$ 时，由 *Brown* 运动的对称性，显然 $P(T_{-a}\leq t)=P(T_a\leq t)$ 所以对于一般的 a ,有

$$P(T_a\leq t)=\sqrt{\frac{2}{\pi}}\int_{\frac{|a|}{\sqrt{t}}}^{+\infty}e^{-\frac{x^2}{2}}dx=2(1-\Phi(\frac{|a|}{\sqrt{t}})) \qquad (2.2)$$

这就是首中时的分布。

由此可以推知两个重要的结论

（1） T_a 几乎处处有限，即 $P(T_a<\infty)=1$. 因为:

$$P(T_a<\infty)=\lim_{t\to\infty}P(T_a\leq t)=\sqrt{\frac{2}{\pi}}\int_0^{+\infty}e^{-\frac{u^2}{2}}du=1 \qquad (2.3)$$

即 a 点首中时 $T_a<\infty$ 的概率为 1。

（2） $ET_a=+\infty$。 (2.4)

证明：

$$ET_a=\int_0^\infty P(T_a>t)dt=\int_0^\infty[1-P(T_a\le t)]dt=\int_0^\infty[1-\sqrt{\frac{2}{\pi}}\int_{\frac{|a|}{\sqrt{t}}}^{+\infty}e^{-\frac{u^2}{2}}du]dt$$

$$=\sqrt{\frac{2}{\pi}}\int_0^\infty\int_0^{\frac{|a|}{\sqrt{t}}}e^{-\frac{u^2}{2}}dudt=\sqrt{\frac{2}{\pi}}\int_0^\infty[\int_0^{\frac{a^2}{u^2}}dt]e^{-\frac{u^2}{2}}du=a^2\cdot\sqrt{\frac{2}{\pi}}\int_0^\infty\frac{1}{u^2}e^{-\frac{u^2}{2}}du$$

$$\ge a^2\cdot\sqrt{\frac{2}{\pi}}\int_0^1\frac{1}{u^2}e^{-\frac{u^2}{2}}du\ge a^2\cdot e^{-1}\sqrt{\frac{2}{\pi}}\int_0^1\frac{1}{u^2}du=+\infty$$

所以 $ET_a=+\infty$。

§3. *Brown* 运动的各种变形与推广

1 吸收 *Brown* 运动

设 $X(t)=\begin{cases}B(t) & T_a>t\\ a & T_a\le t\end{cases}$，其中 $T_a=min\{t:t>0,B(t)=a\}$。

称$\{X(t),t\ge 0\}$为吸收 *Brown* 运动。表示一旦质点到达 a 后，即被吸收停留在 a 点。

可以证明：$X(t)$的概率分布为：

$$\begin{cases}P(X(t)\le y)=1 & \text{当 } y>a\\ P(X(t)=a)=\dfrac{2}{\sqrt{2\pi t}}\displaystyle\int_a^\infty e^{-\frac{u^2}{2t}} & \text{当 } y=a\\ P(X(t)\le y)=\dfrac{2}{\sqrt{2\pi t}}\displaystyle\int_{y-2x}^{y} e^{-\frac{u^2}{2t}} & \text{当 } y<a\end{cases}\tag{3.1}$$

2 反射 *Brown* 运动

设 $Y(t)=|B(t)|$，则称 $\{Y(t),t\ge 0\}$为在原点反射的 *Brown* 运动。

我们来推导 $Y(t)$的分布。显然，当 $y<0$ 时，$P(Y(t)\le y)=0$；当 $y>0$ 时，

$$P(Y(t)\le y)=P(-y\le B(t)\le y)=2P(B(t)\le y)-1=\frac{2}{\sqrt{2\pi t}}\int_{-\infty}^{y}e^{-\frac{u^2}{2t}}du-1\tag{3.2}$$

这样就得到了它的分布。

3 几何 *Brown* 运动

设 $W(t)=e^{B(t)}$，则称$\{W(t),t\geq 0\}$为几何 *Brown* 运动。

几何 *Borwn* 运动有时可以作为相对变化为独立同分布情况的模型。例如，设 $Y_{(n)}$是 n 时刻商品的价格，$\frac{Y_{(n)}}{Y_{(n-1)}}=X_{(n)}$ 是独立同分布的。如取 $Y_{(0)}=1,Y_{(n)}=X_{(1)}X_{(2)}\equiv X_{(n)}$，则

$\ln Y_{(n)}=\sum_{i=1}^{n}\ln X_{(i)}$；所以当 n□∞时，根据中心极限定理知，$\{ln\ Y_{(n)},\ n\geq 1\}$渐近为 *Brown* 运动。于是$\{Y_{(n)},n\geq 0\}$就近似为几何 *Brown* 运动。

取 $B(t)$的矩母函数 $\phi(s)=E[e^{sB(t)}]$，则

$$\phi(s)=E[e^{sB(t)}]=\int_{-\infty}^{\infty}e^{sx}\cdot\frac{1}{\sqrt{2\pi t}}e^{-\frac{x^2}{2t}}dx=e^{\frac{ts^2}{2}},$$

相应地：

$$E(W(t))=E(e^{B(t)})=\phi(1)=e^{\frac{t}{2}}; \tag{3.3}$$

$$D(W(t))=E(W^2(t))-[E(W(t))]^2=E(e^{2B(t)})-e^t=\phi(2)-e^t=e^{2t}-e^t。 \tag{3.4}$$

这样就得到了几何 *Brown* 运动的一阶矩和二阶矩。

4 漂移 *Brown* 运动

设 $\{B(t),\ t\geq 0\}$为 *Brown* 运动，记 $X(t)=B(t)+\mu t$，μ为常数，称$\{X(t),\ t\geq 0\}$ 是带有漂移系数μ的 *Brown* 运动。

漂移 *Brown* 运动的背景是一个质点在直线上作非对称的随机游动，确切叙述如下：一质点在直线上每经 Δt 随机地移动 Δx，每次向右移 Δx 的概率为 p，向左移 Δx 的概率为 q，且每次移动相互独立，以 $X(t)$表示 t 时刻质点位置，令

$$X_i=\begin{cases}1 & 第i次向右移\\ -1 & 第i次向左移\end{cases}，则：X(t)=\Delta x(X_1+X_2+\cdots+X_{[\frac{t}{\Delta t}]}),$$

设 $\Delta x=\sqrt{\Delta t},p=\frac{1}{2}(1+\mu\sqrt{\Delta t}),q=\frac{1}{2}(1-\mu\sqrt{\Delta t})$, 对于给定的$\mu>0$，取充分小的$\Delta t$，使 $\mu\sqrt{\Delta t}<1$；当$\Delta t\to 0$ 时 $EX(t)=\mu t$，$DX(t)=t$，所以 $X(t)\sim N(\mu t,t)$。

将带漂移的 *Brown* 运动的定义写成微分形式，得 $dX(t)=dB(t)+\mu dt$，即质点 t 时刻位移的增量分解为随机性增量与确定性增量之和。

一般地有如下推广：$dX(t)=\sigma dB(t)+\mu dt$。若扩散系数$\sigma$与漂移系数$\mu$不是常数，而是 t 与 $X(t)$ 的函数，那么有如下更一般的随机微分方程：$dX(t)=\sigma(t,X(t))dB(t)+\mu(t,X(t))dt$；这类随机微分方程可用以描述分子的热运动，电子的迁移运动规律等，例如，以 $X(t)$ 描述一个粒子在液体表面 t 时刻的速度，有 $m\dfrac{dX(t)}{dt}=-f\cdot X(t)+\dfrac{dB(t)}{dt}$，其中 m 为质点质量，$-f\cdot X(t)$为粒子与液面的摩擦阻力，$f>0$ 为常数，而 $\dfrac{dB(t)}{dt}$ 为由分子撞击产生的总的合力。求解这一类方程在物理学工程中是常见的，而这离不开 *Brown* 运动理论。

可见漂移 *Brown* 运动很具有实际意义，只要赋予相应系数以物理意义，就可以用它来刻划许多复杂的难以研究的物理过程、工程技术及经济现象。

§4 *Brown* 运动轨道的性质*

本节将通过以下几个命题来研究 *Brown* 运动轨道的性质。

以下均设$\{B(t), t\geq 0\}$ 为标准 *Brown* 运动，且 $B(0)=0$。

命题 4.1 对$\forall t>0$ 固定，有：
$$P\left\{\omega:\lim_{n\to\infty}\sum_{k=1}^{2^n}\left(B\left(\frac{k}{2^n}t\right)-B\left(\frac{k-1}{2^n}t\right)\right)^2=t\right\}=1 \tag{4.1}$$

证明：若 $\forall n\in N$ 记：$X_n=\sum_{k=1}^{2^n}\left(B\left(\frac{k}{2^n}t\right)-B\left(\frac{k-1}{2^n}t\right)\right)^2$，则，只需证：$X_n\xrightarrow{a.s.}t$，

而由第五章定理 4.1 的推论 1，只需证明：$\forall\varepsilon>0$，$\sum_{n=1}^{\infty}P(\omega:|X_n-t|\geq\varepsilon)<\infty$。事实上，我们易知：

$$EX_n=E\left[\sum_{k=1}^{2^n}\left(B\left(\frac{k}{2^n}t\right)-B\left(\frac{k-1}{2^n}t\right)\right)^2\right]=\sum_{k=1}^{2^n}E\left(B\left(\frac{k}{2^n}t\right)-B\left(\frac{k-1}{2^n}t\right)\right)^2=\sum_{k=1}^{2^n}\frac{1}{2^n}t=t$$

故由契比雪夫不等式我们知道：$\forall\varepsilon>0$，$P(\omega:|X_n-t|\geq\varepsilon)\leq\dfrac{DX_n}{\varepsilon^2}$，再令

$Y_k=B\left(\frac{k}{2^n}t\right)-B\left(\frac{k-1}{2^n}t\right)$，则 $X_n=\sum_{k=1}^{2^n}Y_k^{\ 2}$；又 $Y_k\sim N\left(0,\frac{1}{2^n}t\right)$，且 Y_k^2 $(1\leq k\leq n)$ 相互独立，

所以：$DX_n=D\sum_{k=1}^{2^n}Y_k^2=\sum_{k=1}^{2^n}DY_k^2$；再由正态分布的性质知

$$DY_k^2 = EY_k^4 - (EY_k^2)^2 = 3\cdot(\frac{1}{2^n}t)^2 - (\frac{1}{2^n}t)^2 = 2\cdot(\frac{1}{2^n}t)^2 \therefore DX_n = 2\cdot(\frac{1}{2^n}t)^2\cdot 2^n = \frac{1}{2^{n-1}}t$$

故 $\sum_{n=1}^{\infty} P(\omega : |X_n - t| \geq \varepsilon) \leq \sum_{n=1}^{\infty}\frac{DX_n}{\varepsilon^2} = \frac{t}{\varepsilon^2}\sum_{n=1}^{\infty}\frac{1}{2^{n-1}} = \frac{2t}{\varepsilon^2} < \infty$。

命题 4.2 对∀t>0 固定，有：$\underset{\lambda\to 0}{Lim}\sum_{k=1}^{n}(B(t_k)-B(t_{k-1}))^2 \overset{m.s.}{=} t$ (4.2)

证明：要证明以上命题，只需证明 $\underset{\lambda\to 0}{Lim}E[\sum_{k=1}^{n}(B(t_k)-B(t_{k-1}))^2 - t]^2 = 0$。为此计算 $E[\sum_{k=1}^{n}(B(t_k)-B(t_{k-1}))^2 - t]^2$。记 $Y_k = B(t_k)-B(t_{k-1})\ (1\leq k\leq n)$， 故 $Y_k \sim N(0, t_k - t_{k-1})$，且由正态分布的性质知 $EY_k^4 = 3\cdot(t_k - t_{k-1})^2, EY_k^2 = DY_k = t_k - t_{k-1}$, 且 $Y_k^2\ (1\leq k\leq n)$ 相互独立，故当 $k\neq l$ 时，$EY_k^2Y_l^2 = EY_k^2EY_l^2 = (t_k - t_{k-1})(t_l - t_{l-1})$，因此：

$$E[\sum_{k=1}^{n}(B(t_k)-B(t_{k-1}))^2 - t]^2 = E[\sum_{k=1}^{n}Y_k^2 - t]^2 = E(\sum_{k=1}^{n}Y_k^2)^2 - 2E(\sum_{k=1}^{n}Y_k^2)t + t^2$$
$$= 3\cdot\sum_{k=1}^{n}(t_k - t_{k-1})^2 + 2\sum_{i<j}(t_i - t_{i-1})(t_j - t_{j-1}) - 2[\sum_{k=1}^{n}(t_k - t_{k-1})]\cdot t + t^2 = 2\sum_{k=1}^{n}(t_k - t_{k-1})^2$$

$\because \sum_{k=1}^{n}(t_k - t_{k-1})^2 \leq \lambda\cdot\sum_{k=1}^{n}(t_k - t_{k-1}) = \lambda t \ \therefore$ 当 $\lambda\to 0$ 时，$E[\sum_{k=1}^{n}(B(t_k)-B(t_{k-1}))^2 - t]^2 \to 0$

故： $\underset{\lambda\to 0}{Lim}\sum_{k=1}^{n}(B(t_k)-B(t_{k-1}))^2 \overset{m.s.}{=} t$

命题 4.3 对∀t>0 固定，有：$P\left\{\omega : \underset{n\to\infty}{Lim}\sum_{k=1}^{2^n}|B(\frac{k}{2^n}t) - B(\frac{k-1}{2^n}t)| = \infty\right\} = 1$ (4.3)

证明：由命题 4.1 知道：$P(\omega : \underset{n\to\infty}{Lim}\sum_{k=1}^{2^n}(B(\frac{k}{2^n}t) - B(\frac{k-1}{2^n}t))^2 = t) = 1$， 又因为 $B(t)$轨道以概率 1 连续，所以可以证明极大值函数 $\max_{1\leq k\leq 2^n}\{|B(\frac{k}{2^n}t) - B(\frac{k-1}{2^n}t)|\}$ 在闭 区 间 [0,t]

上以概率 1 一致连续，故记：$A=\{\omega:\sum_{k=1}^{2^n}(B(\frac{k}{2^n}t)-B(\frac{k-1}{2^n}t))^2\xrightarrow{a.e.}t\}$，

$B=\{\omega:\max\limits_{1\le k\le 2^n}\{|B(\frac{k}{2^n}t)-B(\frac{k-1}{2^n}t)|\}$一致连续$\}$，则 $P(A)=1, P(B)=1$,又：

$$\sum_{k=1}^{2^n}|B(\frac{k}{2^n}t)-B(\frac{k-1}{2^n}t)|\ge\frac{\sum\limits_{k=1}^{2^n}[B(\frac{k}{2^n}t)-B(\frac{k-1}{2^n}t)]^2}{\max\limits_{1\le k\le 2^n}\{|B(\frac{k}{2^n}t)-B(\frac{k-1}{2^n}t)|\}}$$

对一切ω都成立，而当ω∈AB 且 n□∞时，$\sum_{k=1}^{2^n}[B(\frac{k}{2^n}t)-B(\frac{k-1}{2^n}t)]^2\to t$，且由 $B(t)$的连续

性有：$\max\limits_{1\le k\le 2^n}\{|B(\frac{k}{2^n}t)-B(\frac{k-1}{2^n}t)|\}$ □0，故此时：$\sum_{k=1}^{2^n}|B(\frac{k}{2^n}t)-B(\frac{k-1}{2^n}t)|\to\infty$，而

$P(A)=P(AB)+P(AB^c)$，$0\le P(AB^c)\le P(B^c)=1-P(B)=0$，故：$P(AB)=P(A)=1$。

$$P\left\{\omega:\underset{n\to\infty}{Lim}\sum_{k=1}^{2^n}|B(\frac{k}{2^n}t)-B(\frac{k-1}{2^n}t)|=\infty\right\}=1$$

命题 4.3 说明对 t 在任意区间上，对几乎所有的ω，*Brown* 运动 $B(t,\ \omega)$关于 t 不是有界变差函数。

注：对于任意固定的 t，几乎对所有的ω，$B(t)$关于 t 没有有限的导数。

命题 4.4 对∀t>0 固定，有：

$$\underset{\lambda\to 0}{Lim}\sum_{k=1}^{n}B(t_{k-1})(B(t_k)-B(t_{k-1}))\overset{m.s.}{=}\frac{1}{2}B^2(t)-\frac{1}{2}t \qquad (4.4)$$

$$\underset{\lambda\to 0}{Lim}\sum_{k=1}^{n}B(t_k)(B(t_k)-B(t_{k-1}))\overset{m.s.}{=}\frac{1}{2}B^2(t)+\frac{1}{2}t \qquad (4.5)$$

证明：令 $A_n=\sum_{k=1}^{n}B(t_k)(B(t_k)-B(t_{k-1})),\ C_n=\sum_{k=1}^{n}B(t_{k-1})(B(t_k)-B(t_{k-1}))\ (n\in N)$，则

$$A_n+C_n=\sum_{k=1}^{n}(B(t_k)+B(t_{k-1}))((B(t_k)-B(t_{k-1}))=\sum_{k=1}^{n}(B^2(t_k)-B^2(t_{k-1}))=B^2(t)$$

由命题 4.2 得 $A_n-C_n=\sum_{k=1}^{n}(B(t_k)-B(t_{k-1}))^2\xrightarrow{m.s.}t\ (\lambda\to 0)$ 故：$A_n=B^2(t)-C_n$，而

由均方收敛的定义知： $\underset{\lambda\to 0}{Lim}\, E(A_n - C_n - t)^2 = 0$ $\quad \therefore \underset{\lambda\to 0}{Lim}\, E(B^2(t) - 2\cdot C_n - t)^2 = 0$

$\therefore \underset{n\to\infty}{Lim}\, E[C_n - (\frac{1}{2}B^2(t) - \frac{1}{2}t)]^2 = 0$ 故由此

$$\underset{\lambda\to 0}{Lim}\, E[\sum_{k=1}^{n} B(t_{k-1})(B(t_k) - B(t_{k-1})) - (\frac{1}{2}B^2(t) - \frac{1}{2}t)]^2 = 0，$$

（4.4）式成立。同理可以证明（4.5）式。

在介绍漂移 *Brown* 运动时，我们提到了随机微分方程，与分析中类似，要想解随机微分方程，就必须定义随机积分，也称伊藤积分，然后将相应的微分方程转化为积分方程来求解。关于 *Brown* 运动的积分是最简单也是最常用的一种。

定义 4.1 (伊藤积分) 称一随机过程$\{g(t), t\geq 0\}$关于 *Brown* 运动的伊藤积分是指：对于区间 $[0, t]$ 的任一分割 $0 = t_0 < t_1 < \cdots\cdots < t_n = t$ ，记： $\lambda = \max_{1\leq k\leq n}(t_k - t_{k-1})$ ，若和式 $\sum_{k=1}^{n} g(t_{k-1})(B(t_k) - B(t_{k-1}))$ 当λ□0 时的均方极限存在，则称该极限为 $g(t)$关于 $B(t)$的伊藤积分，并记作：$\int_0^t g(s)dB(s)$。也即

$$\underset{\lambda\to 0}{Lim}\sum_{k=1}^{n} g(t_{k-1})(B(t_k) - B(t_{k-1})) \overset{m.s.}{=} \int_0^t g(s)dB(s) \tag{4.6}$$

例 4.1 求：*Brown* 运动关于自身的伊藤积分$\int_0^t B(s)dB(s)$。

解：由定义知道：$\underset{\lambda\to 0}{Lim}\sum_{k=1}^{n} B(t_{k-1})(B(t_k) - B(t_{k-1})) \overset{m.s.}{=} \int_0^t B(s)dB(s)$， 而由命题 4.4 知：

$$\underset{\lambda\to 0}{Lim}\sum_{k=1}^{n} B(t_{k-1})(B(t_k) - B(t_{k-1})) \overset{m.s.}{=} \frac{1}{2}B^2(t) - \frac{1}{2}t，$$

故由均方收敛极限的唯一性知：

$$\int_0^t B(s)dB(s) = \frac{1}{2}B^2(t) - \frac{1}{2}t \tag{4.7}$$

由上例可以看到，伊藤积分与普通积分（这里指斯蒂尔吉斯的积分，下同），有以下几点明显的不同：

（1） 伊藤积分定义中最后一步取极限是在均方意义下的极限，而不能定义和理解为当

$\delta_n \to 0$ 时，是对每一个ω的求极限。这是因为由命题 4.3 可知，对几乎所有的ω，$B(t,\omega)$ 关于 t 在任意给定小区间$[a,b]$上都不是有界变差函数，故而对几乎所有的ω，其积分和式都不存在有限的极限。

（2） 伊藤积分在定义积分和式时，被积函数在小区间上要取左端点，这与普通积分可任意取不同。这样做的理由可由命题 4.4 看出，若不是固定取左端点，而固定取右端点或任意取，那就会导致极限值不同或极限根本不存在。

（3） 伊藤积分的结果通常不满足普通积分中的牛顿－莱布尼兹公式，以上务必请初学者注意。

伊藤积分和伊藤随机积分方程在现代信息科学、管理科学与金融科学中已成为最重要的数学模型之一，请有兴趣的读者看相关的文献。

练习题

8.1 设 $\{B(t), t \geq 0\}$ 为标准 *B.M.*。(1) $\forall s, t > 0$，求 $E[B(s)B(t)]$；（2）求 $B(t)$ 在 $B(s) = x_1, B(u) = x_2$ 下的 *c.p.d.f.* 及 $E[B(t) \mid B(s) = x_1, B(u) = x_2]$，其中 $0 < s < t < u$；（3）$\forall 0 \leq s < t < u$，求 $E[B(s)B(u) \mid B(t)]$；（4）令 $X(t) = tB(1/t)$，当 $t > 0, X(0) = 0$，当 t=0 时，问若 $0 < t_1 < t_2 < t_3$，$X(t_1)$ 与 $X(t_3) - X(t_2)$ 是否独立？并证之。

8.2 设 $\{B(t), t \geq 0\}$ 为标准 *B.M.*。（1）给定 $B(s)=x$，求 $B(s+t)$ 的条件概率密度；（2）证 $\frac{\partial p}{\partial t} = \frac{1}{2}\frac{\partial^2 p}{\partial x^2}$；（3）$\forall 0 \leq t_1 < t_2 < \cdots < t_{n+1}$，给定 $B(t_i) = x_i, 1 \leq i \leq n$，求 $B(t_{n+1})$ 的条件概率密度，及 $P\{B(t_{n+1}) \leq x \mid B(t_1) = x_1, \cdots, B(t_n) = x_n\}$，$E\{B(t_{n+1}) \mid B(t_1), \cdots, B(t_n)\}$，$Var\{B(t_{n+1}) \mid B(t_1) = x_1, \cdots, B(t_n) = x_n\}, \forall x \in R$

（4）$\forall 0 \leq t_{n+1} < t_n < \cdots < t_1$，给定 $B(t_i) = x_i, 1 \leq i \leq n$。求 $B(t_{n+1})$ 的条件概率密度函数。

8.3 $Y(t) = tB(\frac{1}{t}), Y(0) = 0, Z(t) = B(t) - tB(1), W(t) = \frac{1}{a}B(a^2 t), a > 0, U(t) = (t+1)Z(t/(t+1))$,

(1) 证明：$\{Y(t), t \geq 0\}$，$\{W(t), t \geq 0\}$，$\{U(t), t \geq 0\}$ 都是 *B.M.*；(2)求 $Z(t)$ 的 *p.d.f.*。

8.4 令 $V(t) = \exp(-\alpha t)B[\exp(2\alpha t)]$，求证：$\{V(t), t \geq 0\}$ 是平稳正态过程（称为 *Ornstein-Uhlenback* 过程）并求 $E(V(s)V(t))$ 及 $V(t)$ 的概率密度函数。

8.5 求 $|B(t)|, |\min_{0 \leq s \leq t} B(s)|, M(t) = \max_{0 \leq s \leq t}(B(s))$ 及 $\delta(t) = M(t) - B(t)$ 的概率密度函数。

8.6 求 $S(t) = \int_0^t B(s)ds$ 的协方差、方差及 $(S(t_1)S(t_2))$ 的联合概率密度函数。

8.7 证明 $P\{M(t) > x \mid B(t) = M(t)\} = \exp\left(-x^2/2t\right)$。

提示：求在给定$\delta(t)=M(t)-B(t)=0$下$M(t)$的条件分布。

8.8 设$\mu>0, M=\max_{t\geq 0}X(t)$，证：$\alpha>0, P(M>\alpha)=\exp(2\mu\alpha)$。

8.9 设$n\geq 1, 1\leq k\leq 2^n$，记$\Delta_{nk}=B\left(\frac{k}{2^n}\right)-B\left(\frac{k-1}{2^n}\right), S_n=\sum_{k=1}^{2^n}\Delta_{nk}^2$，（1）求$ES_2$，

$E(S_3\mid S_2),\ E(S_2\mid S_3), ES_n$；（2）证明: $E(S_{n+1}\mid S_n)=\frac{1}{2}(S_n+1)$及

$E(S_n\mid S_{n+1})=S_{n+1}$。

8.10 求$P(T_1<T_{-1}<T_2)$。提示：利用全概公式及*B.M.*的对称性。

8.11 $V(t)$同题 8.4，求(1) $\underset{h\to 0}{Lim}\,E\left[\frac{V(t+h)-V(t)}{h}\mid V(t)\right]$，(2) $\underset{h\to 0}{Lim}\,E\left[\frac{(V(t+h)-V(t))^2}{h}\mid V(t)\right]$

第九章 马尔可夫链

本章讨论离散参数 $T=\{0,1,2,\propto\}=N_0$, 状态空间 $S=\{1,2,\propto\}$可列的马尔可夫过程, 通常称为马尔可夫链(*Markov Chains*), 简称马氏链(*M.C.*)。马氏链最初由 *Markov* 于 1906 年研究而得名。 至今它的理论已发展得较为系统和深入, 它在自然科学, 工程技术及经济管理各领域中都有广泛的应用。

§1 定义与例子

定义 1.1 随机序列$\{X_n, n\geq 0\}$称为马尔可夫链, 如果对任意 $i_0, i_1, \propto, i_n, i_{n+1}\in S, n\in N_0$ 及 $P(X_0=i_0,X_1=i_1,\propto,X_n=i_n)>0$, 有:

$$P(X_{n+1}=i_{n+1}|X_0=i_0, X_1=i_1, \propto, X_n=i_n)=P(X_{n+1}=i_{n+1}|X_n=i_n)。\qquad(1.1)$$

(1.1)式刻划了马氏链的特性, 称为马尔可夫性(或无后效性), 简称马氏性。

定义 1.2 $\forall\, i,j\in S$, 称 $P(X_{n+1}=j|X_n=i)=p_{ij}(n)$为 n 时刻的一步转移概率。 若对 $\forall\, i, j\in S$, $p_{ij}(n)\equiv p_{ij}$ 与 n 无关, 则称$\{X_n, n\geq 0\}$为齐次马氏链。 记 $\boldsymbol{P}=(p_{ij})$, 称 $\boldsymbol{P}$ 为$\{X_n, n\geq 0\}$的一步转移概率矩阵, 简称为转移矩阵(*Transition matrix*)。

本章仅限于讨论齐次马氏链。

为直观理解马氏性的意义, 设想一质点在一直线上的整数点上作随机运动。 以 X_n 表示该质点在时刻 n 的位置。 ($X_n=i$)表示质点在时刻 n 处在 i 状态(位置)这一随机事件。如果把时刻 n 看作“现在”, 时刻 $0, 1, \propto, n-1$ 表示“过去”, 时刻 $n+1$ 表示“将来”, 那么 (1.1)式表明在已知过去 $X_0=i_1, \propto, X_{n-1}=i_{n-1}$ 及现在 $X_n=i_n$ 的条件下, 质点在将来时刻 $n+1$ 处于状态 i_{n+1}(移动到 i_{n+1} 位置)的条件概率, 只依赖于现在发生的事件($X_n=i_n$), 而与过去历史曾发生过什么事件无关。 简言之, 在已知“现在”的条件下, “将来”与“过去”是独立的。$p_{ij}(n)$表示质点在时刻 n 由状态 i 出发, 于时刻 $n+1$ 转移到状态 j 的条件概率。 而齐次性:$p_{ij}(n)=p_{ij}$ 表示此转移概率与时刻 n 无关。

例 1.1 独立随机变量和的序列

设$\{Y_n,n\geq 0\}$为独立同分布随机变量序列。 且 Y_n 取值为非负整数, $P\{Y_n=i\}=a_i$, $a_i\geq 0$, 且$\sum_{i=1}^{\infty}a_i=1$。 令 $X_0=0, X_n=\sum_{i=1}^{n}Y_i$, 则易证$\{X_n, n\geq 0\}$是一马氏链, 且 $p_{ij}=\begin{cases}a_{j-i}, & j\geq i\\ 0, & j<i\end{cases}$

显然$\{Y_n, n\geq 0\}$本身也是一马氏链。

例 1.2 直线上的随机游动

在第七章,我们详细的讨论了一维直线整数点上随机游动的两种最为常见的模型,以下讨论一些更为一般的模型。

(a) 无限制的随机游动

设有一质点在数轴上随机游动, 每隔一单位时间Δt (设 Δt=1)移动一次, 每次只能向左或向右移动Δx 单位(设Δx=1), 或原地不动。 设质点在 0 时刻的位置为 a, 它向右移动的概率为$p\geq 0$ 向左移动的概率为$q\geq 0$, 原地不动的概率为 $r\geq 0$, ($p+q+r=1$), 且各次移动相互独立, 以 X_n 表示质点经 n 次移动后所处的位置。 则$\{X_n, n\geq 0\}$是一马氏链, 且
$p_{i,i+1}=p, p_{i,i-1}=q, p_{ii}=r$, 其余 $p_{ij}=0$。

(b) 带吸收壁的随机游动

设 (a) 中的随机游动限制在$S=\{0, 1, 2, \propto, b\}$内, 当质点移动到状态 0 或$b$后就永远停留在该位置, 即 $p_{00}=1, p_{bb}=1$, 其余 p_{ij} ($1\leq i,j\leq b-1$)同 (a)。 这时序列$\{X_n, n\geq 0\}$称为带两个吸收壁 0 和 b 的随机游动, 是一有限状态马氏链。

(c) 带反射壁的随机游动

如(b)中的质点到达 0 或 b 后, 下一次移动必返回到 1 或 $b-1$, 即 $p_{01}=1, p_{b,b-1}=1, p_{0j}=0$ ($j\neq 1$), $p_{bj}=0$ ($j\neq b-1$), 其余同 (a)。 称此$\{X_n, n\geq 0\}$为带反射壁 0 和 b 的随机游动。 则$\{X_n, n\geq 0\}$是一马氏链。

(d) 艾伦费斯特(*Ehrenfest*)模型

这是一个著名的粒子通过薄膜进行扩散过程的数学模型。 即一质点在状态空间$S=\{-a, -a+1, \propto, -1, 0, 1, 2, \propto, a\}$中作随机游动, 且带有两个反射壁 a 和$-a$, 其一步转移概率是:

$$\begin{cases} p_{i,i-1}=\frac{1}{2}(1+\frac{i}{a}), p_{i,i+1}=\frac{1}{2}(1-\frac{i}{a}), & -a+1\leq i\leq a-1 \\ p_{a,a-1}=1, p_{-a,-a+1}=1, & \\ p_{ij}=0, & \text{其它} \end{cases}$$

由上可看出, 当质点位置 $i<0$,即在原点左边时, $p_{i,i-1}<1/2$, $p_{i,i+1}>1/2$, 此时质点下一步向右移动比向左移动的概率大, 且与离原点的距离成正比。 反之亦然。 当质点在原点时, 向左向右的概率相等。 这样的随机游动可作如下两种解释。

(1) 考虑一容器内有 $2a$ 个粒子作随机游动。 设想一个薄膜(界面)将容器分为相等的左、 右两部分 A 和 B。 如用 X_n 表示 n 时刻 B 内的粒子数与 A 内粒子数之差, 并假定每次移动只有两种可能, 一粒子从左到右或一粒子从右到左, (即同一时刻有两个或两个以上粒子移动的概率为 0, 当Δt□0 时作这种假设是合理的)。则 $\{X_n, n\geq 0\}$可用上述模型来描述。

(2) 设一粒子受一“弹簧力”作用, 在直线上作随机游动, 当粒子偏离原点时, 受到一附加的与偏离距离成正比且指向原点的力的作用, 从而使向原点移动的概率增大。 用 X_n 表示粒子在时刻 n 的位置, 则同样可用上述模型来描述。

例 1.3 排队模型

(a) 离散排队系统

考虑顾客到达一服务台排队等待服务的情况。 若服务台前至少有一顾客等待, 则在一单位时间周期内, 服务员完成一个顾客的服务后, 该顾客立即离去; 若服务台前没有顾客, 则服务员空闲。 在一个服务周期内, 顾客可以到达, 设第 n 个周期到达的顾客数 ξ_n 是一个取值为非负整数的随机变量, 且$\{\xi_n, n\ge1\}$ $i.i.d.$。 在每个周期开始时系统的状态定义为服务台前等待服务的顾客数。 若现在状 态为 i, 则下周期的状态 j 应为

$$j=\begin{cases}(i-1)+\xi, & i\ge 1\\ \xi, & i=0\end{cases}$$

其中ξ为该周期内到达的顾客数。 记第 n 周期开始的顾客数为 X_n, 则 $X_{n+1}=(X_n-1)^++\xi_n$, 这里 $a^+=max(a,0)$。 根据马氏链的定义, 容易证明 $\{X_n, n\ge0\}$是一个马氏链。 若设 $P\{\xi_n=k\}=a_k$, $a_k\ge0$ 且$\sum_{k=0}^{\infty}a_k=1$, 则$\{X_n, n\ge0\}$的转移概率为

$$\begin{cases} p_{0j}=a_j & j\ge 0,\\ p_{1j}=a_j & j\ge 0,\\ p_{ij}=a_{j+1-i} & i>1, j\ge i-1,\\ p_{ij}=0 & i>1, j<i-1,\end{cases}$$

直观上，若 $E\xi_n=\sum_{k=0}^{\infty}ka_k>1$, 则当 n 充分大后, 等待顾客的队长将无限增大; 若 $E\xi_n<1$, 则等待服务的顾客队长趋于某种平衡。

(b) $G/M/1$ 排队系统

G–表示顾客到达服务台的时间间隔, 假设为独立同分布,分布函数为 $G(x)$;

M–表示服务时间, 假设为独立同指数分布(设参数为 μ),且与顾客到达过程相互独立;

1–表示单个服务员。

记 X_n 表示第 n 个顾客到达服务台时系统内的顾客数(包括该顾客), T_n 表示第 n 个顾客到达时刻。 易证 $\{X_n, n\ge1\}$为一马氏链。 为计算它的一步转移概率矩阵，令 $A=\{X_n=i, X_{n+1}=i+1-j\}=\{X_n=i,$ 在(T_n,T_{n+1})上服务完 j 个顾客$\}$，($i\ge0, 0\le j\le i$) 而 $P\{$在$(0,t]$上服务完 j 个顾客$\}=e^{-\mu t}(\mu t)^j/j!$，于是， $P(A\mid X_n=i)=\int_0^{+\infty}P(A\mid X_n=i,T_{n+1}-T_n=t)dG(t)=$

$(j!)^{-1}\int_0^\infty e^{-\mu t}(\mu t)^j\,dG(t)$，得：$p_{i,i+1-j}=(j!)^{-1}\int_0^\infty e^{-\mu t}(\mu t)^j\,dG(t)$，$i\ge0, 0\le j\le i$。而 p_{i0} 是服务员由 i 个顾客转为空闲的概率，易知

$$p_{i,0}=\sum_{k=1}^\infty\int_0^\infty (k!)^{-1}e^{-\mu t}(\mu t)^k\,dG(t)，i\ge0$$

例 1.4 离散分支过程

考虑某一群体, 假定某一代的每一个个体可以产生ξ个下一代个体, 其中ξ是取值为非负整数的离散随机变量, $P\{\xi = k\}=a_k\ge0$, $k\ge0$, $\sum_{k=0}^\infty a_k=1$,设某一代各个体产生下一代的个数相互独立同分布且与上代相互独立。记 X_n 表示第 n 代个体的数目, 则当 $X_n=0$ 时, 有 $X_{n+1}=0$; 当 $X_n>0$ 时, 有 $X_{n+1}=\xi_1+\xi_2+\cdots+\xi_{X_n}$，其中 ξ_i 是第 n 代中第 i 个个体产生的下一代的个数。 由此可知, 只要给定 X_n, 那么 X_{n+1} 的分布就完全决定了, 且与以前的 $X_{n-1}, X_{n-2}, \propto$无关, 故$\{X_n, n\ge1\}$是一马氏链。 其转移概率请读者自己写出。

下面的定理提供了一个非常有用的获得马尔可夫链的方法, 并可用于检验一随机过程是否为马尔可夫链, 读者可以用前面的有关例子验证。

定理 1.1 设随机过程 $\{X_n, n\ge0\}$满足:

1) $X_n=f(X_{n-1}, \xi_n)$ $(n\ge0)$,其中 $f: S\times S\square S$，ξ_n 取值在 S 上;

2) $\{\xi_n, n\ge1\}$为独立同分布随机变量，X_0 与$\{\xi_n, n\ge1\}$也相互独立;

则$\{X_n, n\ge0\}$是马尔可夫链, 其一步转移概率为:$p_{ij}=P(f(i,\xi_1)=j)$。

证明：留给读者作为练习。

§2 转移概率矩阵

设$\{X_n, n\ge0\}$为马氏链, $\boldsymbol{P}=(p_{ij})$, 其中 $p_{ij}=P\{X_{n+1}=j|X_n=i\}$是一步转移概率。 显然

$$p_{ij}\ge0, i,j\in S;\ \sum_{j\in S}p_{ij}=1, i\in S.$$

定义 2.1 称矩阵 $\boldsymbol{A}=(a_{ij})_{S\times S}$为随机矩阵,若$a_{ij}\ge0, i,j\in S$;且对$\forall i\in S$,有 $\sum_{j\in S}p_{ij}=1$.

显然, $\boldsymbol{P}=(p_{ij})$是一随机矩阵。

记

$$\pi_i(n)=P(X_n=i),\ \pi(n)=(\pi_1(n),\pi_2(n),\cdots\pi_i(n)\cdots),$$

$\pi(n)$表示 n 时刻 X_n 的概率分布向量。 称$\{\pi_i(0), i\in S\}$为马氏链的初始分布。 下面, 我们将看到一个马氏链的特性完全由它的一步转移概率矩阵 $\boldsymbol{P}$ 及初始分布向量$\pi(0)$决定。

首先, 对任意 $i_0, i_1, \propto, i_n \in S$, 我们计算有限维联合分布 $P(X_0=i_0, X_1=i_1, \propto, X_n=i_n)$。由概率的乘法公式及马氏性可知

$P(X_0=i_0, X_1=i_1, \propto, X_n=i_n)$

$= P(X_0=i_0)P(X_1=i_1|X_0=i_0)P(X_2=i_2|X_0=i_0, X_1=i_1)\propto P(X_n=i_n|X_0=i_0, \propto,\circ\ X_{n-1}=i_{n-1})$

$= P(X_0=i_0)P(X_1=i_1|X_0=i_0)P(X_2=i_2|X_1=i_1)\propto P(X_n=i_n|X_{n-1}=i_{n-1})$

$$=\pi_{i_0}(0)p_{i_0i_1}p_{i_1i_2}\cdots p_{i_{n-1},i_n}.$$

故 $P(X_0=i_0, X_1=i_1, \propto, X_n=i_n)=\pi_{i_0}(0)p_{i_0i_1}p_{i_1i_2}\cdots p_{i_{n-1},i_n}$，即 $P(X_0=i_0, X_1=i_1, \propto, X_n=i_n)$完全由$\pi(0)$及 $\boldsymbol{P}$ 决定。

类似可以证明任意 n 个时刻的联合分布也完全由 $\pi(0)$及 $\boldsymbol{P}$ 决定。

其次, 有以下定理

定理 2.1

$$\pi(\boldsymbol{n}+1)=\pi(\boldsymbol{n})\boldsymbol{P} \tag{1.2}$$

$$\pi(\boldsymbol{n})=\pi(\boldsymbol{0})\boldsymbol{P}^n \tag{1.3}$$

其中 $\boldsymbol{P}^n$ 是 $\boldsymbol{P}$ 的 n 次幂。

证明: 先对事件进行分解 $(X_{n+1}=j)=\bigcup_{i\in S}(X_n=i, X_{n+1}=j)$, 因为当 $i\neq k$ 时, $(X_n=i)\cup(X_n=k)=\phi$。 故 $P(X_{n+1}=j)=\sum_{i\in S}P(X_n=i, X_{n+1}=j)=\sum_{i\in S}P(X_n=i)P(X_{n+1}=j\,|\,X_n=i)$ $=\sum_{i\in S}\pi_i(n)p_{ij}$ 。 写成向量形式即得 $\pi(\boldsymbol{n}+1)=\pi(\boldsymbol{n})\boldsymbol{P}$ ，重复利用 (1.2) 式即得 (1.3) 式。

(1.3)式表明任一时刻分布由 $\pi(\boldsymbol{0})$ 及 $\boldsymbol{P}$ 完全决定。这些事实表明, 马氏链$\{X_n, n\geq 0\}$的概率性质完全由$\pi(\boldsymbol{0})$与 $\boldsymbol{P}$ 的代数性质决定。

为了下述定理的书写方便, 记 $p_{ij}^{(m)}=P(X_{n+m}=j\,|\,X_n=i)$ 为 m 步转移概率; $\boldsymbol{P}^{(m)}=(p_{ij}^{(m)})$ 为 m 步转移概率矩阵。

定理 2.2 (*Chapman–Kolmogorov* 方程)

$$p_{ij}^{(m+n)}=\sum_{k\in S}p_{ik}^{(m)}p_{kj}^{(n)} \tag{1.4}$$

或

$$\boldsymbol{P}^{(m+n)}=\boldsymbol{P}^{m+n}=\boldsymbol{P}^m\boldsymbol{P}^n=\boldsymbol{P}^{(m)}\boldsymbol{P}^{(n)} \tag{1.5}$$

证明: 因 $(X_0=i, X_{n+m}=j)=\bigcup_{k\in S}(X_0=i, X_m=k, X_{n+m}=j)$ 故:

$$P(X_{n+m}=j\mid X_0=i)=\sum_{k\in S}P(X_0=i,X_m=k,X_{n+m}=j)/P(X_0=i)$$

$$=\sum_{k\in S}P(X_m=k\mid X_0=i)P(X_{n+m}=j\mid X_0=i,X_m=k)$$

$$=\sum_{k\in S}p_{ik}^{(m)}P(X_{n+m}=j\mid X_m=k)\quad（由马氏性）=\sum_{k\in S}p_{ik}^{(m)}\cdot p_{kj}^{(n)}$$

即

$$p_{ij}^{(m+n)}=\sum_{k\in S}p_{ik}^{(m)}p_{kj}^{(n)}$$

写成向量形式即 $\boldsymbol{P}^{(m+n)}=\boldsymbol{P}^{(m)}\boldsymbol{P}^{(n)}$

再注意到 $\boldsymbol{P}^{(1)}=\boldsymbol{P}$, 将 $m=n=1$ 代入上式得 $\boldsymbol{P}^{(2)}=\boldsymbol{P}\ \boldsymbol{P}=\boldsymbol{P}^2$。 从而得到 (1.4) 式及 (1.5) 式。

由上可知, 一个马氏链运动规律的概率特性取决于它的转移概率矩阵特性。 这样, 研究前者就可以转化为研究后者。 (1.4) 式简称 C–K 方程。 显然 $\boldsymbol{P}^{(m)}=(p_{ij}^{(m)})$ 是一随机矩阵。

§3 状态的分类

这一节我们将对马氏链的状态按其概率特性进行分类, 并讨论这些分类的判断准则。

例 3.1 设系统有三种可能状态 $S=\{1,2,3\}$。“1”表示系统运行良好, “2”表示运行正常,“3”表示系统失效。 以 X_n 表示系统在时刻 n 的状态, 并设$\{X_n, n\geq 0\}$是一马氏链。在没有维修及更换条件下, 其自然转移概率为

$$\boldsymbol{P}=\begin{bmatrix}\frac{17}{20} & \frac{2}{20} & \frac{1}{20}\\ 0 & \frac{9}{10} & \frac{1}{10}\\ 0 & 0 & 1\end{bmatrix}$$

由 $\boldsymbol{P}$ 可以看出, 从“1”或“2”状态出发经有限次转移后总要到达“3”状态, 而一旦到达“3”则永远停在“3”。 显然状态“1”, “2”与状态“3”概率性质不同。 由此我们引入如下定义:

定义 3.1 称状态 i 为吸收态 若 $p_{ii}=1$。 对 $i,j\in S$, 若存在 $n\in N$, 使 $p_{ij}^{(n)}>0$, 称自状态 i 出发可达状态 j, 记为 i □j。 如果 i □j 且 j□i, 则称 i,j 相通, 记为 i□j。 若一马氏链的任意两个状态都相通, 则称为不可约链。

定义 3.2 首达时间为:

$$T_{ij}=min\{n: n\geq 1, X_n=j, X_0=i\}。$$

若右边为空集, 则令 $T_{ij}=\infty$。T_{ij}表示从 i 出发首次到达 j 的时间; T_{ii}表示从 i 出发首次回到 i 的时间。

定义 3.3 首达概率为:

$$f_{ij}^{(n)}=P(T_{ij}=n\,|\,X_0=i)=P(X_n=j, X_k\neq j, 1\leq k\leq n-1\,|\,X_0=i)。$$

$f_{ij}^{(n)}$表示从 i 出发经 n 步首次到达 j 的概率。 而 $f_{ij}=\sum_{n=1}^{\infty}f_{ij}^{(n)}$ 表示由 i 出发经有限步终于到达 j 的概率。

定义 3.4 若 $f_{ii}=1$, 称 i 为 常返状态; 若$f_{ii}<1$, 称 i 为 非常返状态 (或称为瞬时态)。

当$f_{ii}=1$ 时, $\{f_{ii}^{(n)}, n\geq 1\}$是一概率分布, 有以下定义:

定义 3.5 如果$f_{ii}=1$, 记 $\mu_i=\sum_{n=1}^{\infty}n\cdot f_{ii}^{(n)}$,则μ_i表示从 i 出发再回到 i 的平均回转时间。若$\mu_i<\infty$, 称 i 为 正常返态; 若$\mu_i=\infty$, 称 i 为 零常返态。

定义 3.6 如果集合$\{n: n\geq 1, p_{ii}^{(n)}>0\}\neq\phi$, 称该数集的最大公约数 $d(i)$为状态 i 的周期。 若 $d(i)>1$ 称 i 为周期的; 若 $d(i)=1$, 称 i 为非周期的。例如在§1 的例 2 (a) 无限制随机游动中, 当 $r=0$, $0<p<1$ 时, $S=\{\propto, -1, 0, 1, 2, \propto\}$, $\{n: n\geq 1, p_{00}^{(n)}>0\}=\{2, 4, 6, \propto\}$。 即状态 0 的 $d(0)=2$, 即从 0 状态出发须经 2 的整数倍次游动才能回到 0 状态, 故它是周期的, 且周期为 2。 当 $p, q, r>0$ 且 $p+q+r=1$ 时, $\{n: n\geq 1, p_{00}^{(n)}>0\}=\{1, 2, 3, \propto\}$, $d(0)=1$, 故此时 0 状态是非周期的。

定义 3.7 若状态 i 为正常返的且非周期的, 则称 i 为 遍历状态。

例 3.2 设马氏链的 $S=\{1, 2, 3, 4\}$, 转移概率矩阵为

$$\boldsymbol{P}=\begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{bmatrix}$$

该链各状态的转移如下图所示

因此 $f_{44}^{(n)}=0, n\geq 1, \therefore f_{44}=0<1; \quad f_{33}^{(1)}=\frac{2}{3}, f_{33}^{(n)}=0, n\geq 2, \therefore f_{33}=\frac{2}{3}<1$

故状态 4 和 3 非常返;由

$$f_{11}=f_{11}^{(1)}+f_{11}^{(2)}=1; \quad f_{22}=\sum_{n=1}^{\infty} f_{22}^{(n)}=0+\frac{1}{2}+\frac{1}{4}+\cdots=1;$$

$$\mu_1=\sum_{n=1}^{\infty} n\cdot f_{11}^{(n)}=1\times\frac{1}{2}+2\times\frac{1}{2}=\frac{3}{2}<\infty\ ;$$

$$\mu_2=\sum_{n=1}^{\infty} n\cdot f_{22}^{(n)}=1\times 0+2\times\frac{1}{2}+\cdots+n\cdot\frac{1}{2^{n-1}}+\cdots=3<\infty\ 。$$

故状态 1 和 2 都是正常返的, 且易知它们是非周期的, 从而是遍历状态。

下面讨论各状态的若干性质以及如何利用转移概率矩阵 $\boldsymbol{P}$ 来判断是否为常返状态。

$p_{ij}^{(n)}$ 与 $f_{ij}^{(n)}$ 有以下关系

定理 3.1 对∀ $i,j\in S$, $n\geq 1$, 有

1)
$$p_{ij}^{(n)}=\sum_{l=1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)}, \tag{3.1}$$

2)
$$f_{ij}^{(n)}=\sum_{k\neq j} p_{ik} f_{kj}^{(n-1)} I_{\{n>1\}}+p_{ij} I_{\{n=1\}}, \tag{3.2}$$

3) i□j□$f_{ij}>0$;
i□j □$f_{ij}>0$, 且 $f_{ji}>0$。

证明: 1)因为

$$\{X_0=i, X_n=j\}=\{X_0=i, X_n=j\}\cap\{\bigcup_{l=1}^{\infty}(T_{ij}=l)\}=\bigcup_{l=1}^{n}(X_0=i, X_n=j, T_{ij}=l)\ 。$$

于是:

$$P(X_0=i)P(X_n=j\mid X_0=i)=\sum_{l=1}^{n} P(X_0=i)P(T_{ij}=l\mid X_0=i)P(X_n=j\mid X_0=i, T_{ij}=l),$$

因此:

$$\begin{aligned}P(X_n=j\mid X_0=i)&=\sum_{l=1}^{n} P(T_{ij}=l\mid X_0=i)P(X_n=j\mid X_0=i, X_k\neq j, 1\leq k\leq l-1, X_l=j),\\&=\sum_{l=1}^{n} f_{ij}^{(l)} P(X_n=j\mid X_l=j) \quad (\text{马氏性})\\&=\sum_{l=1}^{n} f_{ij}^{(l)}\cdot p_{jj}^{(n-l)},\end{aligned}$$

即 :
$$p_{ij}^{(n)}=\sum_{l=1}^{n} f_{ij}^{(l)} p_{jj}^{(n-l)},$$

2)与 3) 的证明与 1)类似,留给读者自己完成。

定理 3.2 状态 i 为常返态, 当且仅当

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty; \tag{3.3}$$

状态 i 为非常返态, 当且仅当

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \frac{1}{1-f_{ii}} < \infty. \tag{3.4}$$

证明: 约定 $p_{ii}^{(0)}=1, f_{ii}^{(0)}=0$. 由(3.1)式有 $p_{ii}^{(n)}=\sum_{l=0}^{n} f_{ii}^{(l)} p_{ii}^{(n-l)}$, 令 $\{p_{ii}^{(n)}\},\{f_{ii}^{(n)}\}(i\in S)$ 的母函数分别为: $P(\rho)$, $F(\rho)$。 即

$$P(\rho)=\sum_{n=0}^{\infty} p_{ii}^{(n)}\rho^n, \qquad F(\rho)=\sum_{n=0}^{\infty} f_{ii}^{(n)}\rho^n.$$

又

$$\sum_{n=1}^{\infty} p_{ii}^{(n)}\rho^n = \sum_{n=1}^{\infty}\left(\sum_{l=0}^{n} f_{ii}^{(l)} p_{ii}^{(n-l)}\right)\rho^n = \left(\sum_{l=1}^{\infty} f_{ii}^{(l)}\rho^l\right)\left(\sum_{n=l}^{\infty} p_{ii}^{(n-l)}\rho^{n-l}\right)$$

$$= (F(\rho)-f_{ii}^{(0)})\sum_{n'=0}^{\infty} p_{ii}^{(n')}\rho^{n'} = F(\rho)\cdot P(\rho) \quad (\because f_{ii}^{(0)}=0)$$

而 $\sum_{n=1}^{\infty} p_{ii}^{(n)}\rho^n = \sum_{n=0}^{\infty} p_{ii}^{(n)}\rho^n - p_{ii}^{(0)}\rho^0 = P(\rho)-1$。因此 $P(\rho)-1=P(\rho)\cdot F(\rho)$.

注意到, 当 $0\le\rho<1$ 时,$F(\rho)<f_{ii}\le 1$, 故

$$P(\rho)=\frac{1}{1-F(\rho)}, \quad 0\le\rho<1 \tag{3.5}$$

又因对一切 $0\le\rho<1$ 及正整数 N, 有

$$\sum_{n=0}^{N} p_{ii}^{(n)}\rho^n \le P(\rho) < \sum_{n=0}^{\infty} p_{ii}^{(n)} \tag{3.6}$$

且当 $\rho\uparrow 1$ 时 $P(\rho)$不减, 故在(3.6) 式中先令 $\rho\uparrow 1$, 后令 $N\to\infty$可得:

$$\lim_{\rho\to 1^-} P(\rho)=\sum_{n=0}^{\infty} p_{ii}^{(n)} \tag{3.7}$$

同理可得:

$$\lim_{\rho\to 1^-} F(\rho)=\sum_{n=0}^{\infty} f_{ii}^{(n)} = f_{ii}. \tag{3.8}$$

于是在(3.5)式中两边令 $\rho\uparrow 1$, 由 (3.7)式和(3.8)式便可得定理的结论。

为解释定理 3.2 的直观意义, 令 $I_n(i)=\begin{cases}1, & X_n=i\\ 0 & X_n\ne i\end{cases}$ 及 $S(i)=\sum_{n=0}^{\infty} I_n(i)$, 则 $S(i)$表示马氏链 $\{X_n, n\ge 0\}$到达 i 的次数。 于是

$$E[S(i)\mid X_0=i]=\sum_{n=0}^{\infty} E[I_n(i)\mid X_0=i]=\sum_{n=0}^{\infty} P(X_n=i\mid X_0=i)=\sum_{n=0}^{\infty} p_{ii}^{(n)} \tag{3.9}$$

可见$\sum_{n=0}^{\infty} p_{ii}^{(n)}$表 示由 i 出发返回到 i 的平均次数。 当 i 为常返态时, 返回 i 的平均次数为无限多次, 反之亦然。 当 i 为非常返态时, 再回到 i 的平均次数至多有限次。

推论 1 若 j 非常返, 则对任意 $i\in S$, 有

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} < \infty, \tag{3.10}$$

$$\lim_{n\to\infty} p_{ij}^{(n)} = 0 \text{。} \tag{3.11}$$

推论 2 若 j 为常返态, 则

1)当 i□j 时, 有

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = \infty \tag{3.12}$$

2) 当 i 不可达 j 时, 有

$$\sum_{n=1}^{\infty} p_{ij}^{(n)} = 0 \tag{3.13}$$

证明：1）当 i□j 时，则存在 $m\geq 1$，使 $p_{ij}^{(m)} > 0$, 于是对 $\forall n\geq 1$, 有 $p_{ij}^{(m+n)} = \sum_{k\in S} p_{ik}^{(m)} p_{kj}^{(n)} \geq p_{ij}^{(m)} p_{jj}^{(n)}$，故 $\sum_{n=1}^{\infty} p_{ij}^{(m+n)} \geq \sum_{n=1}^{\infty} p_{ij}^{(m)} p_{jj}^{(n)} = +\infty$，即得(3.12)式。

2）类似可证(3.13)，留给读者作为练习。

下面再从概率意义考察常返性质。 记:

$$S_m(j) = \sum_{n=m}^{\infty} I_n(j), \quad g_{ij} = P(S_1(j) = +\infty \mid X_0 = i) = P(S_{m+1}(j) = +\infty \mid X_m = i).$$

事件$\{S_m(j) = +\infty\}$表示从时刻 m 起系统无穷多次到达状态 j。 我们有

定理 3.3 对任意 $i\in S$, 有

$$g_{ij} = \begin{cases} f_{ij}, & j\text{常返}, \\ 0, & j\text{非常返}. \end{cases} \tag{3.14}$$

证明: 因 $\{S_m(j)\geq k+1\} \subset \{S_m(j)\geq k\}$, 故

$$(S_1(j) = +\infty) = \bigcap_{k=1}^{\infty} (S_1(j)\geq k) = \lim_{k\to\infty} (S_1(j)\geq k).$$

由概率的连续性可得

$$g_{ij}=P(S_1(j)=+\infty\,|\,X_0=i)=P(\bigcap_{k=1}^{\infty}(S_1(j)\ge k)\,|\,X_0=i).=\lim_{n\to\infty}P(S_1(j)\ge k\,|\,X_0=i).\quad(3.15)$$

又 $(S_1(j)\ge k+1,X_0=i)=\bigcup_{l=1}^{\infty}(T_{ij}=l,S_1(j)\ge k+1)=\bigcup_{l=1}^{\infty}(T_{ij}=l,S_{l+1}(j)\ge k)$ 。故

$$\begin{aligned}P(S_1(j)\ge k+1\,|\,X_0=i)&=\sum_{l=1}^{\infty}P(T_{ij}=l\,|\,X_0=i)P(S_{l+1}(j)\ge k\,|\,X_0=i,T_{ij}=l)\\&=\sum_{l=1}^{\infty}f_{ij}^{(l)}P(S_{l+1}(j)\ge k\,|\,X_0=i,X_m\ne j,1\le m\le l-1,X_l=j)\\&=\sum_{l=1}^{\infty}f_{ij}^{(l)}P(S_{l+1}(j)\ge k\,|\,X_l=j)\quad\text{(马氏性)}\\&=\sum_{l=1}^{\infty}f_{ij}^{(l)}P(S_1(j)\ge k\,|\,X_0=j)\quad\text{(时齐性)}\end{aligned}$$

即 $$P(S_1(j)\ge k+1\,|\,X_0=i)=f_{ij}P(S_1(j)\ge k\,|\,X_0=i)\tag{3.16}$$

反复利用上式可得 $$P(S_1(j)\ge k+1,X_0=i)=f_{ij}(f_{jj})^k\tag{3.17}$$

令 k□∞, 若 j 常返, 即 f_{jj}=1, 则由 (3.15)式及(3.17)式得 $g_{ij}=f_{ij}$, 若 j 非常返, 即 $f_{jj}<1$, 则 g_{ij}=0。

定理 3.4 状态 i 常返, 当且仅当 g_{ii}=1; 若状态 i 非常返, 则 g_{ii}=0。

证明: 将 (3.14) 式中 j 换成 i 即可得。

定理 3.4 的说明: 若 i 常返, 则系统从 i 出发以概率 1 无穷多次返回 i, 即从 i 出发的几乎所有样本轨道无穷多次回到 i; 若 i 非常返, 则从 i 出发几乎所有样本轨道至多有限次回到 i。

进一步地, 若 i 为常返态, 如何判别它是零常返的还是遍历的? 我们先叙述一个重要定理而不加证明。

定理 3.5 设 i 常返且周期为 d, 则

$$\lim_{n\to\infty}p_{ii}^{(nd)}=\frac{d}{\mu_i}.\tag{3.18}$$

其中μ_i为 i 的平均回转时间。 当$\mu_i=+\infty$ 时, 理解为 $d/\mu_i=0$。

定理的证明可参见[11]。

定理 3.6 设 i 常返, 则

1) i 为零常返, 当且仅当 $\lim_{n\to\infty}p_{ii}^{(n)}=0$;

2) i 为遍历, 当且仅当 $\lim_{n\to\infty}p_{ii}^{(n)}=\dfrac{1}{\mu_i}>0$。 (3.19)

证明: 由定理 3.5 容易得到,详细的推导留给读者。

状态相通关系为等价关系, 因为具有

1) 自反性: i□i 。 这由下面定义可得: $p_{ij}^{(0)}=\delta_{ij}$

2) 对称性: 若 i□j, 则 j□i。

3) 传递性: 若 i□j 且 j□k, 则 i□k。

传递性的证明如下：

由于 i□j, j□k, 则 $\exists m,n,$ 使 $p_{ij}^{(n)}>0, p_{jk}^{(m)}>0,$ 则由 $C\text{–}K$ 方程

$$p_{ik}^{(m+n)}=\sum_{l\in S}p_{il}^{(m)}p_{lk}^{(n)}\geq p_{ij}^{(m)}p_{jk}^{(n)}>0.$$

故 i□k 同理可证 k□i, 故 i□k。

利用等价关系, 可以把马氏链的状态空间分为若干等价类。 在同一等价类内的状态彼此相通，在不同的 等价类中的状态不可能彼此相通。 然而, 从某一类出发以正的概率到达另一类的情形是可能的。

定义 3.8 如一马氏链的所有状态属于同一等价类, 则称它是不可约链。

例 3.3

(a)

$$S=\{1,2,3\},\qquad \boldsymbol{P}=\begin{bmatrix}\frac{1}{2}&\frac{1}{4}&\frac{1}{4}\\ \frac{1}{4}&0&\frac{3}{4}\\ 0&\frac{2}{3}&\frac{1}{3}\end{bmatrix}$$

(b)

$$S=\{1,2,3,4,5\},\qquad \boldsymbol{P}=\begin{bmatrix}\frac{1}{2}&\frac{1}{2}&0&0&0\\ \frac{1}{4}&\frac{3}{4}&0&0&0\\ 0&0&0&1&0\\ 0&0&\frac{1}{2}&0&\frac{1}{2}\\ 0&0&0&1&0\end{bmatrix}$$

(c)

$$S=\{1,2,3,4,5\},\qquad \boldsymbol{P}=\begin{bmatrix}0.6&0.1&0&0.3&0\\ 0.2&0.5&0.1&0.2&0\\ 0.2&0.2&0.4&0.1&0.1\\ 0&0&0&1&0\\ 0&0&0&0&1\end{bmatrix}$$

读者可以自己画状态转移图进行判断。

(a) 由于所有状态相通，组成一等价类。 故该链是不可约链。

(b) 此链可分为两个等价类 {1, 2}及{ 3, 4, 5}。

(c) 此链可分为三个等价类{1, 2, 3},{4}及{5}。 由{1, 2, 3}可进入{4} 或{5}，反之则不行。

对于相通的状态，我们不加证明的给出以下重要的定理

定理 3.7 若 i□j，则

1） i 与 j 同为常返或非常返。 若为常返，则它们同为正常返或同为零常返;

2） i 与 j 或有相同的周期，或同为非周期。

该定理说明，对相通的状态，因是同类型，故只需选出其中之一较容易判别的状态即可。

关于常返态的判断，我们可以总结为以下重要定理

定理 3.8 下列命题等价

1) i 为常返态;

2) $P(\bigcup_{n=1}^{\infty}(X_n=i)\,|\,X_0=i)=1;$

3) $P(S_1(i)=+\infty\,|X_0=i)=1;$

4) $\sum_{n=0}^{\infty}p_{ii}^{(n)}=\infty;$

5) $E\{S_1(i)\,|X_0=i\}=+\infty$。

证明: 1) □2): 由于 $f_{ii}=P(T_{ii}<\infty|\,X_0=i)=1$，及

$$(T_{ii}<\infty)=\bigcup_{n=1}^{\infty}(T_{ii}=n)=\bigcup_{n=1}^{\infty}(X_0=i,X_l\neq i,0<l<n,X_n=i)=\bigcup_{n=1}^{\infty}(X_0=i,X_n=i)$$

因此 $$P(\bigcup_{n=1}^{\infty}(X_n=i)\,|\,X_0=i)=P(T_{ii}<\infty\,|\,X_0=i)=f_{ii}=1.$$

即 1)与 2)等价。 1) □3) :见定理 3.4； 1) □4): 见定理 3.2； 4) □5): 由 (3.11)式即得。

例 3.4 (直线上的无限制随机游动) 在例 1.2 (a)中我们已经知道它是马氏链,且容易证明它的全体状态构成一个类,事实上,它其中的任意两个状态都相通。问题是它是常返类还是非常返类? 由定理 3.7 我们知道只需选一个“代表”即可。

下面我们来计算 $\sum_{n=0}^{\infty}p_{ii}^{(n)}$,据此来判别 i 是否常返。由第七章的知识我们知道：

$$p_{ii}^{(2n+1)}=0,\quad p_{ii}^{(2n)}=C_{2n}^{n}p^{n}q^{n}\ (n\in N_0)$$

我们利用定理 3.2 中的母函数 $P(\rho)$ 有

$$P(\rho)=\sum_{n=0}^{\infty}C_{2n}^{n}p^{n}q^{n}\rho^{2n}=\sum_{n=0}^{\infty}\frac{(2n)!}{n!n!}(pq\rho^{2})^{n}=\sum_{n=0}^{\infty}\frac{2^{2n}(-1)^{n}}{n!}(-\frac{1}{2})(-\frac{3}{2})\cdots(-\frac{2n-1}{2})(pq\rho^{2})^{n}$$

$$=\sum_{n=0}^{\infty}\frac{1}{n!}(-\frac{1}{2})(-\frac{3}{2})\cdots(-\frac{2n-1}{2})(-4pq\rho^{2})^{n}=(1-4pq\rho^{2})^{-\frac{1}{2}}。$$

从而有 $\sum_{n=0}^{\infty}p_{ii}^{(n)}=P(1)=\lim_{\rho\to1-0}P(\rho)=\lim_{\rho\to1-0}(1-4pq\rho^{2})^{-\frac{1}{2}}$，故

$$\sum_{n=0}^{\infty}p_{ii}^{(n)}=\begin{cases}\infty, & p=\frac{1}{2}\\ \text{有限}, & p\neq\frac{1}{2}\end{cases}$$

由此知,p=1/2 时,是常返的,从而全体状态构成单一的类是常返的（这样的马是链称为常返链）；p≠1/2 时,链是非常返的。

§4　极限性态与平稳分布

在实际应用中, 人们常常关心的问题有两个: (a) 当 n□∞时, $P(X_n=i)=\pi_i(n)$ 的极限是否存在?(b) 在什么条件下，一个马氏链是一个平稳序列?对于前者,由于 $\pi_j(n)=\sum_{i\in S}\pi_i(0)p_{ij}^{(n)}$，故可转化为研究 $p_{ij}^{(n)}$ 的 渐近性质, 即 $\lim_{n\to\infty}p_{ij}^{(n)}$ 是否存在? 若存在,其极限是否与 i 有关? 对于后者, 实际上是一个平稳分布是否存在的问题。 这两个问题有密切联系。

1　P^n 的极限性态

$p_{ij}^{(n)}$ 的渐近性态, 在§3 中已有所涉及, 这里分两种情形再加以进一步讨论。

(1)　j 非常返或零常返

定理 4.1 若 j 非常返或零常返, 则对任意 $i\in S$, 有 $\lim_{n\to\infty}p_{ij}^{(n)}=0$　　(4.1)

证明: 当 j 为非常返时, 上述结论已在§3 定理 3.2 的推论 1 中证过。故只需证 j 为零常返的情形。 取 $m<n$, 有

$$p_{ij}^{(n)}=\sum_{l=1}^{n}f_{ij}^{(l)}p_{jj}^{(n-l)}\le\sum_{l=1}^{m}f_{ij}^{(l)}p_{jj}^{(n-l)}+\sum_{l=m+1}^{n}f_{ij}^{(l)}.\qquad(4.2)$$

固定 m 先令 n□∞, 由§3 定理 3.6 知, 上式右方第一项趋于 0 (因为 $p_{jj}^{(n)} \to 0$, 且是有限项的和) ; 再令 m□∞ 第二项因 $\sum_{l=1}^{\infty} f_{ij}^{(l)} \le 1$ 而趋于 0。 故 $p_{ij}^{(n)} \to 0 \quad n \to \infty$ 。

(2) j 为正常返态

这时情况较复杂, $\underset{n\to\infty}{Lim}\, p_{ij}^{(n)}$ 不一定存在, 即使存在也可能与 i 有关。但有以下结论:

定理 4.2 若 j 正常返, 周期为 d, 则对任意 $i\in S$ 及 $0\le r\le d-1$, 有

$$\underset{n\to\infty}{Lim}\, p_{ij}^{(nd+r)} = f_{ij}(r)\cdot\frac{d}{\mu_j} \tag{4.3}$$

其中 $f_{ij}(r)=\sum_{m=0}^{\infty} f_{ij}^{(md+r)}, 0\le r\le d-1$ 。

此定理的证明可见参考书[11]。

推论 1 若 j 是遍历状态, 则对任意的 $i\in S$, 有

$$\underset{n\to\infty}{Lim}\, p_{ij}^{(n)} = \frac{f_{ij}}{\mu_j} \tag{4.4}$$

推论 2 对于不可约的遍历链 (即所有状态遍历), 对任意 $i, j\in S$, 有

$$\underset{n\to\infty}{Lim}\, p_{ij}^{(n)} = \frac{1}{\mu_j} \tag{4.5}$$

定理 4.3 若马氏链是不可约的遍历链, 则$\{\pi_i=\frac{1}{\mu_i}\}$是方程组

$$x_j=\sum_{i\in S} x_i p_{ij} \tag{4.6}$$

满足条件$x_j\ge 0, j\in S, \sum_{j\in S} x_j =1$的唯一解。

证明: 记 $\pi_i=\frac{1}{\mu_i}$, 由定理 4.2 推论 2 有 $\underset{n\to\infty}{Lim}\, p_{ij}^{(n)}=\frac{1}{\mu_j}=\pi_j$. 对任意 n, M, 有:

$1=\sum_{j\in S} p_{ij}^{(n)} \ge \sum_{j=1}^{M} p_{ij}^{(n)}$ 固定 M, 令 n□∞, 可得$\sum_{j=1}^{M}\pi_j\le 1$; 再令 M□∞, 知$\sum_{j\in S}\pi_j\le 1$ 。

由 C–K 方程有 $p_{ij}^{(n+1)} \ge \sum_{l=10}^{M} p_{il}^{(n)} p_{lj}$, 如令 n□∞, 再令 M□∞, 有

$$\pi_j \geq \sum_{l\in S}\pi_l p_{lj}, \quad \forall j\in S \tag{4.7}$$

将(4.7)式两边乘以 p_{ji} 并对 j 求和，得 $\pi_i \geq \sum_{j\in S}\pi_j p_{ji} \geq \sum_{l\in S}\pi_l p_{li}^{(2)}$， 重复上述步骤，得：

$\pi_j \geq \sum_{i\in S}\pi_i p_{ij}^{(n)}$， 对所有 $j\in S$ 及 $n\geq 1$ 成立。 现设上式对某个 j 严格不等式成立，即：

$\pi_j > \sum_{i\in S}\pi_i p_{ij}^{(n)}$， 将此式对 j 求和，有 $\sum_{j\in S}\pi_j > \sum_{j\in S}\sum_{i\in S}\pi_i p_{ij}^{(n)} = \sum_{i\in S}\pi_i \sum_{j\in S} p_{ij}^{(n)} = \sum_{i\in S}\pi_i$，

但这是不可能的。 于是对所有 n 及 j，有

$$\pi_j = \sum_{i\in S}\pi_i p_{ij}^{(n)} \tag{4.8}$$

由于 $\sum_{j\in S}\pi_j \leq 1$，且 $p_{ij}^{(n)}$ 关于 n 一致有界，故在(4.8) 式中令 n□∞ 时，由控制收敛定理，有

$$\pi_j = \lim_{n\to\infty}\sum_{i\in S}\pi_i p_{ij}^{(n)} = \sum_{i\in S}\pi_i \cdot \lim_{n\to\infty} p_{ij}^{(n)} = (\sum_{i\in S}\pi_i)\cdot \pi_j .$$

由于 $\pi_j>0$，故得: $\sum_{i\in S}\pi_i = 1$。

现证唯一性。 设 $\{v_i\}$ 是满足条件的另一组解，则类似 (4.8) 有

$$v_j = \sum_{i\in S} v_i p_{ij} = \cdots = \sum_{i\in S} v_i p_{ij}^{(n)} .$$

令 n□∞ 则有 $v_j = \sum_{i\in S} v_i \cdot \lim_{n\to\infty} p_{ij}^{(n)} = (\sum_{i\in S} v_i)\cdot \pi_j = \pi_j$。

2 平稳分布

定义 4.1 一个定义在 S 上的概率分布 $\pi=\{\pi_1,\pi_2, \propto,\pi_i ,\propto\}$ 称为马氏链的平稳分布，如有

$$\pi = \pi \boldsymbol{P} \tag{4.9}$$

即 $\forall j \in S$,

$$\pi_j = \sum_{i\in S}\pi_i p_{ij} . \tag{4.10}$$

平稳分布也称马氏链的不变概率测度。 对于一个平稳分布 π，显然有

$$\pi=\pi \boldsymbol{P} =\pi \boldsymbol{P}^2= \propto=\pi \boldsymbol{P}^n 。 \tag{4.11}$$

定理 4.4 设 $\{X_n,n\geq 0\}$ 是马氏链， 则 $\{X_n,n\geq 0\}$ 为平稳过程的充要条件是 $\pi(0)=(\pi_i (0), i\in S)$ 是平稳分布，即: $\pi (0)=\pi (0) \boldsymbol{P}$。

证明: 充分性。 记 $\pi (0) = \pi$，显然 $\pi (1) =\pi (0) \boldsymbol{P} =\pi$, ∝ $\pi (n)= \pi (n-1) \boldsymbol{P} =\pi \boldsymbol{P} =\pi$。因此 $\forall i_k\in S, t_k\in N, n\geq 1, 1\leq k\leq n, t\in N$ 有

$$P(X_{t_1}=i_1,X_{t_2}=i_2,\cdots X_{t_n}=i_n)=\pi_{i_1}(t_1)p_{i_1i_2}^{(t_2-t_1)}\cdots p_{i_{n-1}i_n}^{(t_n-t_{n-1})}$$
$$=\pi_{i_1}(t_1+t)p_{i_1i_2}^{(t_2-t_1)}\cdots p_{i_{n-1}i_n}^{(t_n-t_{n-1})}$$
$$=P(X_{t_1+t}=i_1,X_{t_2+t}=i_2,\cdots X_{t_n+t}=i_n)。$$

所以$\{X_n, n\geq 0\}$是严平稳过程。

必要性。 由于$\{X_n, n\geq 0\}$是平稳过程，因此有$\pi(n)=\pi(n-1)=\propto=\pi(0)$。又由 $\pi(1)=\pi(0)\boldsymbol{P}$ 得$\pi(0)=\pi(0)\boldsymbol{P}$，即$\pi(0)$ 是平稳分布。

由定理 4.3 有以下结论：

定理 4.5 不可约遍历链恒有唯一的平稳分布$\{\pi_i=\dfrac{1}{\mu_i}\}$，且 $\pi_j=\underset{n\to\infty}{Lim}\,p_{ij}^{(n)}$。

对于一般的马氏链，其平稳分布是否存在？若存在，是否唯一？这里不深入讨论，有兴趣的读者可参看[11]，[14]。

3 $\underset{n\to\infty}{Lim}\,\pi_j(n)$的存在性

下面我们来研究$\underset{n\to\infty}{Lim}\,\pi_j(n)$的存在性问题。

定义 4.2 若$\underset{n\to\infty}{Lim}\,\pi_j(n)=\pi_j^*$ $(j\in S)$ 存在，则称 $\pi^*=\{\pi_1^*,\cdots,\pi_j^*,\cdots\}$为马氏链的极限分布。

定理 4.6 非周期不可约链是正常返的充要条件是它存在平稳分布，且此时平稳分布就是极限分布。

证明：充分性。 设存在平稳分布$\pi=\{\pi_1,\propto,\pi_j,\propto\}$，由此有$\pi=\pi\boldsymbol{P}=\pi\boldsymbol{P}^2=\propto=\pi\boldsymbol{P}^n$ 即 $\pi_j=\sum_{i\in S}\pi_i p_{ij}^{(n)}$，由于$x_j\geq 0, j\in S, \sum_{j\in S}x_j=1$，当 n□∞，利用控制收敛定理，极限号与和式可交换，得$\pi_j=\underset{n\to\infty}{Lim}\sum_{i\in S}\pi_i p_{ij}^{(n)}=\sum_{i\in S}\pi_i\cdot\underset{n\to\infty}{Lim}\,p_{ij}^{(n)}=(\sum_{i\in S}\pi_i)\cdot\pi_j$。因为 $\sum_{j\in S}\pi_j=\sum_{j\in S}\dfrac{1}{\mu_j}=1$。于是至少存在一个$\pi_l=\dfrac{1}{\mu_l}>0$，从而 $\underset{n\to\infty}{Lim}\,p_{il}^{(n)}=\dfrac{1}{\mu_l}>0$ 即$\mu_l<\infty$，故l为正常返状态。 由不可约性知，整个链是正常返的，且所有$\pi_j=\dfrac{1}{\mu_j}>0$。

必要性。 由于马氏链是正常返非周期链，即为遍历链，由定理 4.6 立即得证。 且所有$\pi_j=\pi_j^*=\dfrac{1}{\mu_j}$，$j\in S$。

由上可知，对于不可约遍历链，则极限分布 $\pi^*=\pi$ 存在且等于平稳分布。 这意味

着当 n 充分大时，$P(X_n = j) \approx \pi_j = \dfrac{1}{\mu_j}$，即$\{X_n, n\geq 0\}$是一渐近平稳序列。这在实际问题中是很有意义的。

例 4.1 设

$$S=\{1, 2\}, \boldsymbol{P} = \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{5}{8} & \frac{3}{8} \end{bmatrix}$$

求:平稳分布 及 $\lim\limits_{n\to\infty} \boldsymbol{P}^n$=?

解: 由 π=π$\boldsymbol{P}$ 得 $\pi_1 = \frac{3}{4}\pi_1 + \frac{5}{8}\pi_2$ 及 $\pi_1 + \pi_2 = 1$,解得 $\pi_1 = \frac{5}{7}$，$\pi_2 = \frac{2}{7}$,故: $\pi = (\frac{5}{7}, \frac{2}{7})$

由 $\lim\limits_{n\to\infty} p_{ij}^{(n)} = \pi_j = \dfrac{1}{\mu_j}$，故 $\mu_1 = \dfrac{7}{5}, \mu_2 = \dfrac{7}{2}$ 且

$$\lim_{n\to\infty} \boldsymbol{P}^n = \lim_{n\to\infty} \begin{bmatrix} \frac{3}{4} & \frac{1}{4} \\ \frac{5}{8} & \frac{3}{8} \end{bmatrix}^n = \begin{bmatrix} \frac{5}{7} & \frac{2}{7} \\ \frac{5}{7} & \frac{2}{7} \end{bmatrix}.$$

4 求和(积分)与极限交换的原则

下面罗列几个关于求和(积分)与极限交换的重要定理，这些定理可以看成是实变函数理论中有关定理的推广。

定理 4.7 (*Levy* 单调收敛定理) 设π $=\{\pi_i, i\in S\}$是行向量, $\pi_i \geq 0$, $\forall\ i\in S$, 若列向量序列 $\{f^{(n)}\}, f^{(n)} = (f_1^{(n)}, f_2^{(n)}, \cdots f_i^{(n)}, \cdots)^T$， 满足 $0 \leq f^{(1)} \leq \cdots \leq f^{(n)} \leq f^{(n+1)} \leq \cdots$ 且 $\lim\limits_{n\to\infty} f^{(n)} = f$. 则 $\pi f = \lim\limits_{n\to\infty} \pi f^{(n)}$ 即 $\sum\limits_{i\in S} \pi_i \cdot (\lim\limits_{n\to\infty} f^{(n)}) = \lim\limits_{n\to\infty} \sum\limits_{i\in S} \pi_i f_i^{(n)}$.

定理 4.8 (*Fatou* 定理) 设π $=\{\pi_i, i\in S\}$是行向量，$\pi_i \geq 0$, $\forall\ i\in S$, 若列向量序列 $\{f^{(n)}\}, f^{(n)} = (f_1^{(n)}, f_2^{(n)}, \cdots f_i^{(n)}, \cdots)^T$,满足 $\pi \varliminf\limits_{n\to\infty} f^{(n)} \leq \varliminf\limits_{n\to\infty} \pi f^{(n)}$ 且 $\lim\limits_{n\to\infty} f^{(n)} \doteq f$.存在。

则 $\pi f \leq \varliminf\limits_{n\to\infty} \pi f^{(n)}$.

定理 4.9 (*Lebesgue* 控制收敛定理)设π $=\{\pi_i, i\in S\}$是行向量, $\pi_i \geq 0$, $\forall\ i\in S$, 若列向量序列 $\{f^{(n)}\}, f^{(n)} = (f_1^{(n)}, f_2^{(n)}, \cdots f_i^{(n)}, \cdots)^T$,满足 $f^{(n)}\geq 0$,

$|f^{(n)}| = (|f_1^{(n)}|, |f_2^{(n)}|, \cdots |f_i^{(n)}|, \cdots)^T \leq ce$, $e=(1,1,\propto)^T$, $c>0$ 为常数，且 $\lim\limits_{n\to\infty} f^{(n)} \doteq f$.存在。

则 $\pi f = \lim\limits_{n\to\infty} \pi f^{(n)}$。

§5 离散时间的 *Phase−Type* 分布

本节讨论离散时间的 *Phase−Type* 分布,以下先给出它的定义。

定义 5.1 设$\{X_n, n\geq0\}$是 *M.C.*, $\widetilde{S} = S\bigcup S_0$,其中,$S=\{1, 2,\propto, m\}$为瞬时态集,$S_0=\{0\}$为吸收态 且 $\widetilde{P} = \begin{bmatrix} P & P_0 \\ 0 & 1 \end{bmatrix}$,其中 $P_0=(I-P)e$, $e=(1,\propto,1)^T$, $\tau = \inf\{n : n \geq 0, X_n \in S_0\}$,则称τ为首达时间（或 *PH* 分布）。

令$\pi(0)= (\alpha_0, \alpha)$,其中 $\alpha = (\alpha_1, \propto,\alpha_m)$, $\alpha_k\geq0$, $\sum\limits_{k\in\widetilde{S}}\alpha_k = 1$. $g_k = P(\tau=k)$,

$$G(\lambda)=E(\lambda^\tau) = \sum_{k=0}^{\infty} g_k \lambda^k \tag{5.1}$$

本节的核心问题就是求τ的分布 g_k,则有如下的定理:

定理 5.1 在上述记号下,有

1）$\forall k \in N$, $g_0 = \alpha_0$,

$$g_k = \alpha P^{k-1}P_0 = \alpha P^{k-1}(I-P)e; \tag{5.2}$$

2) $\forall\ 0\leq\lambda\leq1$, $G(\lambda)= \alpha_0 +\lambda\, \alpha\, (I-\lambda P)^{-1} (I-P)e$ (5.3)

证明：1) 用数学归纳法:

k=0 时, $g_0 = P(\tau=0)=P(X_0\in S_0)=\alpha_0$,

k=1 时,$g_1 = P(\tau=1)=P(X_0\in S, X_1=0)=\sum\limits_{i\in S}\alpha_i p_{i0} = \alpha \cdot P_0$,

k=2 时, $g_2 = P(\tau = 2) = P(X_0 \in S, X_1 \in S, X_2 \in S_0)$

$$= \sum_{i\in S}\sum_{j\in S} P(X_0 = i, X_1 = j, X_2 = 0) = \sum_{i\in S}\sum_{j\in S}\alpha_i p_{ij} p_{j0} = \alpha \cdot P^{2-1} \cdot P_0$$

假设 k=n 时命题成立,即 $g_n = \alpha P^{n-1}P_0 = \alpha P^{n-1}(I-P)e$, 而当 k=n+1 时：可仿上作如下的事件分解：1）从初始状态 i 转移一步到 j,2）以 j 作为初始状态然后转移 n 步被吸收,则结合归纳假设有 $g_{n+1} = P(\tau=n+1)= \alpha P\bullet P^{n-1}P_0 = \alpha P^nP_0 = \alpha P^{(n+1)-1}P_0 = \alpha P^n (I-P)e$。知当 k=n+1 时,命题成立。

综上有：$\forall k \in N$, $g_0 = \alpha_0$, $g_k = \alpha P^{k-1}P_0 = \alpha P^{k-1}(I-P)e$。

为证明 2）我们先给出一个引理

引理 设矩阵 Q 满足 $\lim\limits_{n\to\infty} Q^n = 0$,则 $(I-Q)^{-1}$ 存在,且

$$(I-Q)^{-1} = \sum_{k=1}^{\infty} Q^k. \tag{5.4}$$

其中 I 为单位矩阵。

证明：因

$$(I-Q)(I+Q+\cdots+Q^{n-1}) = (I-Q^n). \tag{5.5}$$

若 $\lim\limits_{n\to\infty} Q^n = 0$，则当 n 充分大时，$\det(I-Q^n) = |I-Q^n| \neq 0$，从而：

$$|I-Q| \cdot |I+Q+\cdots+Q^{n-1}| \neq 0$$

于是 $|I-Q| \neq 0$，得 $(I-Q)^{-1}$ 存在，再以 $(I-Q)^{-1}$ 左乘(5.5)式两边，并让 $n\to\infty$ 即得(5.4)式。

以下证明 2)

将 g_k 的表达式代入(5.1)式中，并注意到 S 为瞬时态集，故 $\lim\limits_{n\to\infty} P^n = 0$,用引理可得 2)

PH 分布有如下的一些性质

性质 5.1 若τ_1,τ_2 为 *PH* 分布,则$\tau_1+\tau_2$, $\tau_1\wedge\tau_2$, $\tau_1\vee\tau_2$ 均为 *PH* 分布。

性质 5.2 τ_1, $\propto\tau_n$ 为 *PH* 分布,ξ为 *d.r.v.*且与τ_1, $\propto\tau_n$ 独立,$P(\xi=k) = p_k$, $1\le k\le n$，则：

$\sum\limits_{k=1}^{n} \tau_k I_{(\xi=k)}$ 为 *PH* 分布。

练习题

9.1 抛掷一枚硬币，以 H_n 和 T_n 分别记前 n 次抛掷中“上”和“下”的次数，令 $X_n = H_n, Y_n = H_n - T_n$,它们是马氏链吗？如果是，求转移矩阵。

9.2 设 S={1,2,3}, 三个马氏链的转移矩阵分别为 ：

$$P_1 = \begin{bmatrix} 1 & 0 & 0 \\ \frac{1}{3} & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}, P_2 = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{2} & 0 & \frac{2}{3} \\ 0 & \frac{1}{3} & \frac{2}{3} \end{bmatrix}, P_3 = \begin{bmatrix} 0 & \frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix}$$

（1）试分别对状态进行分类;

（2）对 P_3 求 T_{11} 的分布律及 ET_{11}。

9.3 一个国家在稳定经济条件下它的出口商品能够用三状态的马氏链描述如下: 状态空间 S = { +1, 0, -1 }； +1: 今年比去年增长；0：波动低于； -1: 今年比去年减少。由以往的统计数据求得转移矩阵为

$$\begin{array}{c} \\ +1 \\ 0 \\ -1 \end{array}\begin{array}{c} \begin{array}{ccc} +1 & 0 & -1 \end{array} \\ \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.35 & 0.3 & 0.35 \\ 0 & 0.4 & 0.6 \end{bmatrix} \end{array}$$

试求每个状态的平均返回时间, 并比较在稳定经济条件下增长趋势与减少趋势的期望长度。

9.4 水库供水按其水位分为下列 5 个状态: “1”-- 危险水平; “2”-- 缺水; “3”-- 刚够; “4”-- 较好; “5”-- 充裕。S= { 1, 2, 3, 4, 5 }, 由已有数据求得相邻时间周期的转移矩阵为

$$P = \begin{bmatrix} 0.1 & 0.1 & 0.3 & 0.5 & 0 \\ 0.3 & 0.2 & 0.2 & 0.2 & 0.1 \\ 0.1 & 0.2 & 0.4 & 0.2 & 0.1 \\ 0 & 0.1 & 0.2 & 0.4 & 0.3 \\ 0 & 0.1 & 0.1 & 0.4 & 0.4 \end{bmatrix}$$

试求出现危险水平的平均时间长度(即求 $\mu_{11} = \mu_1$)。

9.5 设马氏链 $\{X_n, n \ge 0\}$, S= { 1, 2, 3 }, P 如下 ：

$$P=\begin{bmatrix}0.5 & 0.4 & 0.1\\ 0.3 & 0.4 & 0.3\\ 0.2 & 0.3 & 0.5\end{bmatrix},$$

（1）求平稳分布$\pi=(\pi_1,\pi_2,\pi_3)$及 $\lim_{n\to\infty}P^n$；

（2）当初始分布$\pi(0)$是怎样分布时，此马氏链是平稳序列？并求EX_n及DX_n。

9.6 设 S= { 1, 2, 3 }, $P=\begin{bmatrix}\frac{1}{2} & \frac{1}{4} & \frac{1}{4}\\ 0 & \frac{3}{4} & \frac{1}{4}\\ 0 & 0 & 1\end{bmatrix}$， 求：（1）$T_{13}$的分布律及$ET_{13}$；

（2）$f_{ii}(i=1,2,3)$；（3）$n\to\infty$时$P^n\to ?$。

9.7 一马氏链$\{X_n,n\ge 0\}$的 S= { 0, 1, 2, …，N}，转移概率为

$$P_{ij}=\begin{cases}\mu_i, & j=i-1\\ \lambda_i, & j=i+1\\ 1-\lambda_i-\mu_i, & j=i\\ 0, & |j-i|>1\end{cases}\qquad i,j\in S,$$

$\mu_0=\lambda_0=\mu_N=\lambda_N=0, 0<\mu_i<0, 0<\lambda_i<1, 1\le i\le N-1, X_0=k$,

求：(1) P（$X_n=0$，某个$n\ge 0\mid X_0=k$）；

(2) P（$X_n=N$，某个$n\ge 0\mid X_0=k$）；（3）ET_{k0}及ET_{kN}。

9.8 设 j 为非常返状态，证明对任意$i\in S$，有 $\sum_{n=1}^{\infty}p_{ij}^{(n)}=f_{ij}\big/(1-f_{jj})<\infty$。

9.9 设有两串独立的贝努里(*Bernoulli*)试验序列$X_1,X_2,\cdots$及$Y_1,Y_2,\cdots$，它们成功的概率分别为 p_1 与 p_2，即：$P(X_i=1)=1-P(X_i=0)=p_1, P(Y_i=1)=1-P(Y_i=0)=p_2$，$\{X_n,n\ge 0\}$与$\{Y_n,n\ge 0\}$独立。为决定是否 $p_1>p_2$ 或 $p_2\ge p_1$，我们利用下列检验：选取某个正整数 M 使得：

或者 $X_1+X_2+\cdots+X_n-(Y_1+Y_2+\cdots+Y_n)=M$，

或者 $X_1+X_2+\cdots+X_n-(Y_1+Y_2+\cdots+Y_n)=-M$ 。若试验结果是前一情况发生，则判定 $p_1>p_2$；若是后一情况发生，则判定 $p_2\ge p_1$。记

$N=\min\{n:n\ge 1, X_1+X_2+\cdots+X_n-(Y_1+Y_2+\cdots+Y_n)=\pm M\}$ 。

试证明：（1）在 $p_1>p_2$ 条件下，经试验误判 $p_2>p_1$ 的概率为 $1/(1+\lambda^M)$，其中 $\lambda=\dfrac{p_1(1-p_2)}{p_2(1-p_1)}$；（2）$EN=M(\lambda^M-1)/(p_1-p_2)(\lambda^M+1)$。

9.10 设 $\{Y_n,n\ge 0\}$ *i.i.d.,* $P(Y_n=1)=p\ge 0, P(Y_n=-1)=q=1-p\ge 0$，

令 $X_0=Y_0=i$，$X_N=\sum_{k=1}^{n}Y_k, n\ge 1, T_i=\min\{n:n\ge 0, X_0=i,\ X_n=0或X_n=b\}$

其中 b 为正整数，$0\le i\le b$。（1）当 i=1, b=3 时，求 $P(T_1=k\mid X_0=1), k\in N$ 及 $P(X_{T_1}=3\mid X_0=1)$；（2）当 i=2, b=5 时，求 $P(T_2=k\mid X_0=2), k\in N$ 及 $P(X_{T_2}=5\mid X_0=2)$。

9.11 设从数字 1，2，…，a 中任取一数为 $X_0, n\ge 1$ 时，则从 1，2，…，X_{n-1} 中任取一数为 X_n，则 $\{X_n,n\ge 0\}$ 为马尔可夫链，试写出它的转移概率矩阵。

9.12 设连续的掷一个均匀的硬币，定义 $X_n=0$，若第 n 次掷得正面；$X_n=k$，若第 *n-1*，…，*n-k+1* 次掷得反面，而第 *n-k* 次掷得正面。这里我们规定第 0 次掷得正面，则 $\{X_n,n\ge 0\}$ 为马尔可夫链，试写出它的转移概率矩阵。

9.13 一个“传染病”，模型如下：设 a 个人中某些人已患流行性感冒。假定：（1）当一个病人遇见一个健康者时，后者被传染的概率为 α；（2）所有的接触都是两人之间的接触；（3）一切成对的接触都是等可能的；（4）在每个单位时间内只发生一次接触。以 X_n 记时刻 n 患病的人数，则 $\{X_1,\cdots,X_n\}$ 为马尔可夫链，试写出它的转移概率矩阵。

9.14 考察汽车单向通过某路口的情形。假设一秒钟可通过一辆汽车。路口设置有自动红绿灯，红灯持续 r 秒，绿灯持续 g 秒。以“红灯-绿灯”为一个单元时间，在第 n 个单元开始时在等候的汽车数记为 X_{n-1}，在这单元中来到的汽车数记为 ξ_n。设

$\{\xi_n, n \geq 1\}$ 为 *i.i.d.*随机变量,其共同分布为 $\{q_k, k=0,1,\cdots,r+g\}$,而开始时路口没有汽车在等候,则 $\{X_n, n \geq 0\}$ 为马尔可夫链，试写出它的转移概率矩阵。

9.15 设某一水库的容量为 c 立方米.以 X_n 记时刻 n 水库的储水量,在时间区间[n,n+1)内流入水库的水量为 ξ_{n+1} ，超过水库容量的水即溢出，在时间区间[n,n+1）末从水库放掉 m（<c）立方米水（其中包括库满而溢出的部分，如果溢出的水量已超过 m 立方米就不再放水）。假设 $\{\xi_n, n \geq 1\}$ *i.i.d.* 且与 X_0 独立， $P(\xi_n = k) = \alpha_k, k \geq 0$ ，则 $\{X_n, n \geq 0\}$ 为马尔可夫链，试写出它的转移概率矩阵。

9.16 设 $\{X_n, n \geq 0\}$ 为（p，q）—随机游动，0<q<p<1，初始分布为：

$P(X_0 = 0) = P(X_0 = 2) = 1/2$ 。取 $A = \{1,3\}$ ，证明：

$$P(X_2 = 2 \mid X_0 = 0, X_1 \in A) \neq P(X_2 = 2 \mid X_1 \in A),$$

这表明：对马尔可夫链 $\{X_n, n \geq 0\}$ ，若状态集合 A 包含不止一个状态，则下列等式一般不成立（即使 A 只有两个状态）：

$$P(X_{n+1} = j \mid X_0 = i_0, \cdots, X_{n-1} = i_{n-1}, X_n \in A) = P(X_{n+1} = j \mid X_n \in A)。$$

9.17 设 $\{\xi_n, n \geq 1\}$ *i.i.d.*, $P(\xi_n = 1) = P(\xi_n = -1) = 1/2$, $S_0 = 0, S_n = \sum_{k=1}^{n} \xi_k, n \geq 1;$

$M_n = \max\limits_{0 \leq k \leq n} S_k, n \geq 0$ 。证明： $\{|S_n|, n \geq 0\}$ 及 $\{M_n - S_n, n \geq 0\}$ 为马尔可夫链。

9.18 设有限马尔可夫链有 a 个状态，k 为常返状态。证明：存在常数 q，0<q<1，使得 n>a 时，从 k 出发首次返回 k 的时间大于 n 的概率 $P_k(\tau_k > n) < q^n$ 。

9.19 设 j 为非常返状态， N_j 为 $\{X_n, n \geq 0\}$ 中状态 j 出现的次数。证明：

$$P(N_j = r) = \begin{cases} f_{ij}(1-f_{jj})f_{jj}^{r-1}, & r \geq 1 \\ 1 - f_{ij}, & r = 0 \end{cases}。$$

9.20 设 τ 为 X 的有限停时，令 $Y = \{Y_n, n \geq 0\}, Y_n = X_{\tau+n}, n \geq 0$ 。证明：Y 为马尔可夫链，与 X 有相同的转移概率矩阵，且在已知 $Y_0 = X_\tau$ 的条件下，Y 与 $(\tau, X_0, \cdots, X_\tau)$

条件独立。

9.21 设转移概率为：$p_{00}=1, p_{10}=(1-\delta)q, p_{11}=\delta q, p_{12}=p, p_{ii}=q$，

$p_{i.i+1}=p, i\ge 2, 0<p, \delta<1, p+q=1$。求吸收概率 $f_{i0}, i\ge 1$。

9.22 在马尔可夫链中定义 $P_{ij}(z)=\sum_{n=0}^{\infty}p_{ij}^{(n)}z^n$， $F_{ij}(z)=\sum_{n=0}^{\infty}f_{ij}^{(n)}z^n, |z|<1$。

证明：对一切 $t, j\in E$， $P_{ij}(z)=\delta_{ij}+F_{ij}(z)P_{ij}(z), |z|<1$。由此可得：

$P_{ii}(z)=1\big/(1-F_{ii}(z)), |z|<1$，令 $z\to 1$, 由于 $F_{ii}(1)=f_{ii}$, 我们又得到常返判别法：

$f_{ii}=1$ 等价于 $\sum\limits_{n=0}^{\infty}p_{ii}^{(n)}=\infty$ 的一个证明。

9.23 在（p，q）—随机游动中，证明：$f_{00}^{(2n)}=\begin{cases}2pq, & n=1\\ \frac{1}{2}\frac{3}{2}\cdots\frac{2n-3}{2}\frac{(4pq)^n}{2(n!)}, & n\ge 2\end{cases}$。

9.24 设 $\{\xi_n, n\ge 1\}$ *i.i.d.*， $P(\xi_n=k)=\alpha_k, k\ge 1$，对 $n\ge 0$ 定义：

$Y_n=\begin{cases}0, & n<\xi_1\\ k, & \xi_1+\cdots+\xi_k\le n<\xi_1+\cdots+\xi_{k+1}, k\ge 1\end{cases}$， $X_n=n-\sum_{j=0}^{Y_n}\xi_j, (\xi_0=0)$，

则 $\{X_n, n\ge 0\}$ 为常返不可约马尔可夫链，正常返的充要条件为 $\sum_{k=1}^{\infty}ka_k<\infty$。

9.25 设 $\{X_n, n\ge 0\}$ 为遍历不可约马尔可夫链，转移概率矩阵为 $P=(P_{ij})$， $\{\xi_n, n\ge 1\}$ 为 *i.i.d.*随机变量，其分布为 $P(\xi_n=k)=\alpha_k, k\ge 1$，且 $\{X_n, n\ge 0\}$ 与 $\{\xi_n, n\ge 1\}$ 相互独立。令 $\eta_0=0, \eta_n=\xi_1+\cdots+\xi_n\ n\ge 1$， $Y_n=X_{\eta_n}, n\ge 0$，则 $\{Y_n, n\ge 0\}$ 为正常返不可约马尔可夫链，其转移概率矩阵为 $\tilde{P}=\sum_{k=1}^{\infty}a_kP^k$ ，且与 $\{X_n, n\ge 0\}$ 有相同的平稳分布。

第十章 参数估计

§1 数理统计的研究对象及基本概念

通过前几章的介绍可知,很多随机现象的规律性可以用随机变量和随机过程来描述,而要刻划它们需要知道相应的概率分布,或它们的某些数字特征和参数。然而,在许多实际问题中对这些往往知之甚少,或只知部分信息。对此,用什么方法才能确定该随机变量的分布或其参数呢?这就需要掌握数理统计的基本理论与方法。

1 数理统计的研究对象

从本章起,简要介绍数理统计的两个基本内容:估计与检验。数理统计是随机数学的重要组成部分,它和概率论联系非常密切,可以说:概率论是数理统计的重要基础,而数理统计是概率论的重要应用。本书力求对蕴含在数理统计中的那些独特的思维方式与推断依据作一些粗浅介绍,并给出一些典型的例子,而不去罗列一大堆各种不同形式的具体做法。务请读者重点放在领略思路及举一反三能力的训练上。

数理统计应用范围极为广范,至今已渗透到人类社会生活的各个领域。例如,在一大堆杂乱无章的信息与数据面前,如何分析处理,去伪存真,找出其数据中反映的规律与模型。又如要研制一种新的药品(或新产品),而影响药品的功能与质量的因素非常繁多(不仅有成分的配比还有工艺条件等),如何全面与科学地安排试验,以便在较短时间内找出疗效好的药品配方与工艺条件,如何找出影响性能的主要因素以及它们的数量关系?又如,一个工厂每天生产一大批产品,如何进行质量管理与控制,如何预测未来产品的销售量?又如,面对天花乱坠的股票波动与金融风险,如何作出较有把握的预测,抓住机遇作出明智的决策?等等。数理统计将为这些问题提供富有启发性的思维方法与强有力的工具。下面仅举几个具体例子:

(1)如何估计产品的寿命?这是工业品质量管理中极重要的问题。为了评价某批电子设备的使用寿命,随机抽取了 18 台做试验,测得寿命数据如下(单位:小时):17,29,50,68,100,130,140,270,280,340,410,450,520,620,190,210,800,1100。
假设使用寿命服从指数分布。问:整批电子设备中,寿命超过 200 小时的占多大比例?(由本章随后要介绍的参数估计方面的知识可知,这个比例大约是 $e^{-200/318}=0.533$。)

(2)为了探讨吸烟与患慢性支气管炎是否关联,调查了 339 人,情况如下表:

人数	患慢性支气管炎	未患慢性支气管炎	合计
吸烟	43	162	205
不吸烟	13	121	134
合计	56	283	339

问：从这批数据能否断定患慢性支气管炎与吸烟有关？

（答案是肯定的。 这是一个统计推断问题）

数理统计的主要内容是：应用概率论和数学方法，研究怎样收集（通过试验或观察）带有随机误差的信息与数据，建立适当的随机数学模型，并在设定的模型（称为统计模型）之下，对这种信息与数据进行分析（称为统计分析）。以对所研究的问题作出推断（称为统计推断）与决策。下面再举一例来说明数理统计研究的二个基本内容：

例 1.1 某工厂每月生产大批电子元件。按以往的试验数据分析结果认为元件的寿命服从指数分布，即 $f(x)=\lambda e^{-\lambda x}I_{(x>0)}$，但每月生产元件的参数 λ 并不知道，那么：（1）该月生产元件的平均寿命如何估计？ （2）如果你是使用单位，要求平均寿命能达到某个指定的数，例如 5000 小时。问这批元件可否被接受？

如果该分布中的参数 λ 已知，则可知其平均寿命为 $1/\lambda$，于是上面两个问题马上就可以得到回答。但现在 λ 是未知的，于是须要从一大批元件中随机抽出若干个，例如 n 个，并测出其寿命分别为 $X_1,\cdots,X_n$。然后将得到的部分信息，经过一定的加工，便能较好地估计未知参数 λ 与平均寿命。例如用其算术平均值 $\overline{X}=(X_1+\cdots+X_n)/n$ 去估计未知的平均寿命 $1/\lambda$。

问题（1）称为参数估计问题。这时考虑使用 $1/\overline{X}$ 作为未知参数 λ 的估计。显然由于随机抽取，自然 $\overline{X}$ 亦是随机的，那么（a）这种估计它的精度与误差有多大？（b）估计量与 λ 之差不超过给定的允许误差有多大把握（概率可靠度）？（c）为要保证满足给定的“精度”和“可靠度”，至少要多大样本容量 n 才能达到目的？（d）有没有更好的估计方法？等等这些都是参数估计研究的主要内容。参数估计是数理统计中的基本问题之一。

问题（2）与（1）不同：问题（2）是要判断总体的均值是否达到 5000。通常把对总体的某种要求作为对总体的一个假设（记为 H_0）提出，然后用样本的观察结果对总体的假设 H_0 进行检验，最后作出接受或拒绝 H_0 的决择。显而易见，当 $\overline{X}$ 取值 $\bar{x}$ 比 5000 大很多，就应该接受 H_0，且 $\overline{X}$ 越大就越有理由与把握接受 H_0；反之，就应拒绝 H_0。那么（a）对给定的假定 H_0，如何制定检验 H_0 成立的检验准则？（b）给定一检验法，通过检验，接受与拒绝 H_0 的概率各有多大？（c）如何比较与衡量不同检验法的优劣？（d）为了减少推断失误的概率，该选择多大的 n 才能达到目的？等等这类问题称为假设检验

问题，也是最重要的统计问题之一。

2 总体和样本

总体是指所研究的对象全体构成的集合。总体的元素，称为个体。如在例 1.1 中一大批元件就是问题的总体，而每一单个元件就是一个个体。在数理统计中，主要关心的是对象的某些数量指标的统计规律性。因此可用随机变量 X 的取值全体作为总体，简称 X 为总体（或母体）。X 的全部统计规律性可用 X 的分布函数 $F(x)$来刻划，称 $F(x)$为总体分布，记为 $X\sim F(x)$。有时，总体 X 的分布也用 $f(x)$表示，当总体 X 为离散型随机变量（简记为 *d.r.v.*）时，它表示分布律；当总体 X 为连续型随机变量（简记为 *c.r.v.*）时，它表示概率密度函数（简称密度函数，简记为 *p.d.f.*）。

为了研究总体的性质，似乎最好是把每个个体的指标都加以观测，但这往往是不必要的，有时甚至是不可能的，怎么办呢？一个主要的方法，即从总体中选取少量个体作为代表，即随机抽样法。然后对这抽取的部分个体进行观测与研究，从而推断总体的性质。从总体中抽取的部分个体称为样本。从总体 X 中抽取的 n 个个体记为 $X_1,\cdots,X_n$，这 n 个个体 $X_1,\cdots,X_n$ 称为总体 X 的一个容量为 n 的样本（或称子样）。本书上的样本通常是按“简单抽样”抽取一部分个体。所谓“简单抽样”是指总体中的每个个体每次被抽出是等可能的。为了使样本更具代表性，且能从有限的样本中包含与提取更多的信息，要求抽取（观测）是彼此独立的且与总体同分布。由此引进下面的定义：

定义 1.1 称 $X_1,\cdots,X_n$ 是来自总体 X 的一个容量为 n 的样本（子样），若 $X_1,\cdots,X_n$ *i.i.d.*，且每个 X_i 均与 X 同分布。n 称为样本容量，$X_1,\cdots,X_n$ 的全体称为一组样本，而 $X_i\,(1\le i\le n)$ 称为其中的第 i 个样本。

显然，若总体 X 的分布为 $f(x)$，则其样本 $X_1,\cdots,X_n$ 的分布为：$\prod_{k=1}^{n} f(x_k)$。

3 统计量

$X_1,\cdots,X_n$ 作为来自总体 X 的一组样本，包含了总体分布的信息，我们要利用它对 X 的分布规律进行统计推断，但因为样本是 n 维的，这些信息是分散到样本的每个分量上的。因此，直接从样本出发来推断总体分布是不方便的。为此，需要对样本进行加工，将样本中分散的信息通过适当的变换，浓缩起来，以便更加突出地显示出总体某一侧面的特性。所谓统计量，就是不含未知参数的样本的函数。特别地，统计量只依赖于样本，而不依赖未知参数。我们要根据不同的需要，对样本进行不同的加工与提炼，从而把样本中蕴含的总体的不同侧面的特性突出反映出来，就可以对总体的某些参数与性质作出

尽可能精度高的估计与可靠的推断。

例如，设 $X_1,\cdots,X_n$ 是从正态总体 $N(\mu,\sigma^2)$ 中抽出的样本，则 $\overline{X}=(X_1+\cdots+X_n)/n$ 是统计量，因为它完全由样本 $X_1,\cdots,X_n$ 决定。而当参数 μ 未知时，$\overline{X}-\mu$ 就不是统计量。同理，$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 也是统计量。通常称 $\overline{X}$ 为样本均值，称 S^2 为样本方差。显然统计量 $\overline{X}$ 与 S^2 是对样本的不同加工，它们各自显露出总体不同侧面（例如 μ与σ^2）的性态。常用的统计量是样本均值与样本方差等。

定义 1.2 设总体 $X\sim F(x)$, $X_1,X_2,\cdots,X_n$ 为其样本，$EX=\theta, DX=\sigma^2$, 则分别称

$$\overline{X}=\sum_{i=1}^{n}X_i/n \quad , \quad S^2=\sum_{i=1}^{n}(X_i-\overline{X})^2\Big/(n-1)$$

为样本均值与样本方差，称

$$\sum_{i=1}^{n}X_i^k/n \quad , \quad \sum_{i=1}^{n}(X_i-\overline{X})^k/n$$

为样本 k 阶原点矩与样本 k 阶中心矩。

命题 1.1 $E\overline{X}=\theta, D\overline{X}=\sigma^2/n, ES^2=\sigma^2$. （1.1）

证明：留给读者作为练习。

统计量的分布称为抽样分布。其中常用的分布有正态分布，卡方分布 $\chi^2(n)$，t 分布 $t(n)$，F 分布 $F(m,n)$。下面简要给出后三种分布的定义、密度函数及其一些性质。

1. 自由度为 *n* 的卡方分布：

定义 1.3 设 $X_1,\cdots,X_n$ *i.i.d.*，$X_1\sim N(0,1^2)$，则称随机变量 $\chi^2(n)=\sum_{i=1}^{n}X_i^2$ 的分布是自由度为 n 的卡方分布，简记为 $\chi^2(n)$。即 $\chi^2(n)$ 分布就是相互独立的标准正态分布的 n 个随机变量平方和的分布。

命题 1.2 $\chi^2(n)$ 分布的密度函数为：

$$K_n(x)=\frac{1}{\Gamma(n/2)2^{n/2}}e^{-x/2}x^{(n-2)/2} \qquad (1.2)$$

其中 $\Gamma(\alpha)$ 为伽玛函数。

证明梗概：先求 X_1^2 的 $p.d.f.$，（见第二章例题），然后利用和的卷积公式（见第三章）与数学归纳法即得。有兴趣的读者可作为练习补上。

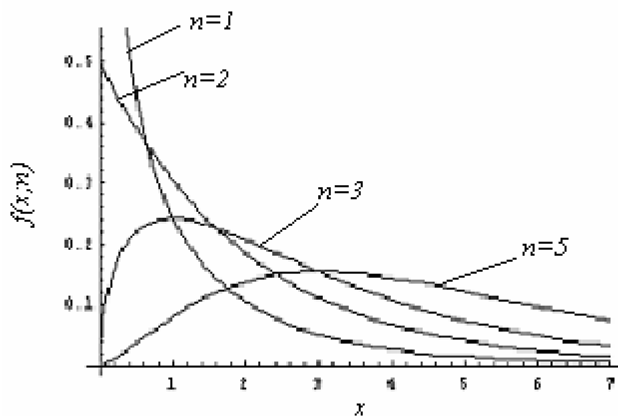


图 1.1

$\chi^2(n)$ 分布的密度函数如图 1.1。

卡方分布有如下重要性质：

(1) 设 X_1, X_2 独立，$X_1 \sim \chi^2(m)$，$X_2 \sim \chi^2(n)$，则 $X_1+X_2 \sim \chi^2(m+n)$。

证明：设 $Y_1,\cdots, Y_m, Y_{m+1}, \cdots\ Y_{m+n}$ $i.i.d.$且 $Y_i \sim N(0,1)$。令 $X_1=\sum_{i=1}^{m} Y_i^2$，$X_2=\sum_{j=1}^{n} Y_{m+j}^2$，则 $X_1 \sim \chi^2(m)$，$X_2 \sim \chi^2(n)$，且 X_1, X_2 独立（见第三章 $r.v.$独立性质），$X_1+X_2=\sum_{i=1}^{m+n} Y_i^2$ 为 $m+n$ 个标准正态 $r.v.$的平方和。按 χ^2 分布的定义知其分布为 $\chi^2(m+n)$，即得证。

(2) 若 $X_1, \cdots, X_n$ 独立且都服从指数分布 $f(x)=\lambda e^{-\lambda x} I_{(x>0)}$，则

$$X = 2\lambda \sum_{i=1}^{n} X_i \sim \chi^2(2n)$$

证明：首先，由 X_i 的 $p.d.f.$可推出 $2\lambda X_i$ 的 $p.d.f.$为 $\frac{1}{2} e^{-x/2} I_{(x>0)}$。但由（1.1）当 n=2 时,可知这正好是 $\chi^2(2)$ 的 $p.d.f.$，即 $2\lambda X_i \sim \chi^2(2)$。再因 $X_1, \cdots, X_n$ 独立，利用刚才证明的性质，即得所要的结果。

2. 自由度为 *n* 的 *t* 分布：

定义 1.4 设 X_1,X_2 独立，X_1～$N(0,1)$， X_2～$\chi^2(n)$，则称

$$T(n)=X_1\Big/\sqrt{X_2/n}$$

的分布是自由度为 n 的 t 分布,简记为 $t(n)$，亦称为学生（*student*）分布。这种分布是英国人 *w.s.Gosset* 在 1908 年以笔名”*student*”发表的，它是数理统计中最重要的分布之一。

命题 1.3 设 $T(n)$ 是自由度为 n 的 t 分布，则它的 *p.d.f.*为：

$$t_n(y)=\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\,\Gamma(n/2)}(1+y^2/2)^{-(n+1)/2} \tag{1.3}$$

证明梗概：为证(1.3)，由 X_2 的 *p.d.f.*先求 $\sqrt{X_2/n}$ 的 *p.d.f.*，再利用第三章推广的全概率公式，求 *r.v.*商 $X_1\Big/\sqrt{X_2/n}$ 的 *p.d.f.*，命题可得证。

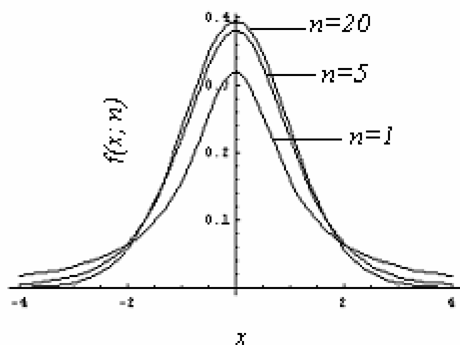


图 1.2

这个密度函数关于原点对称，其图形与标准正态分布 $N(0,1)$的密度函数的图形类似。见图 1.2。由中心极限定理易知：当 $n\to\infty$ 时，t 一分布的极限分布是标准正态分布。

3. *F* 分布：

定义 1.5 设 X_1,X_2 独立，X_1～$\chi^2(n)$，X_2～$\chi^2(m)$，则称

$$F=m^{-1}X_2\Big/n^{-1}X_1$$

的分布是自由度为(m, n)的 F 分布，简记为 $F(m, n)$。

命题 1.4 设 $F=m^{-1}X_2\Big/n^{-1}X_1$ 如上定义，则 F 的 *p.d.f.*为：

$$f_{m,n}(y)=m^{m/2}n^{n/2}\,\frac{\Gamma((m+n)/2)}{\Gamma(m/2)\Gamma(n/2)}\,y^{m/2-1}(my+n)^{-(m+n)/2}I_{(y>0)} \tag{1.4}$$

证明：直接由定义应用推广的全概率公式于 *r.v.*商的情形，即可证得(1.4)。

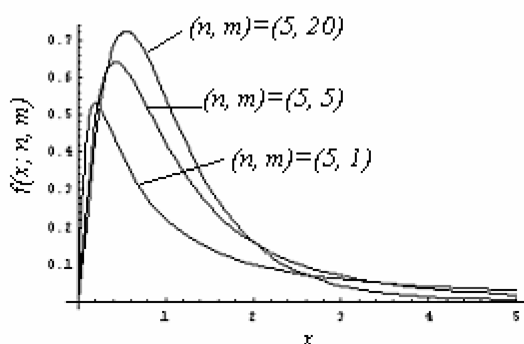


图 1.3

$F(m, n)$分布的图形见图 1.3。

性质：如 $X\sim F(m, n)$，则 $1/X\sim F(n, m)$。

通常把χ^2，t 和 F 这三个分布合称“统计上的三大分布”，是因为它们在统计学中有广泛的应用。

下面给出正态总体的样本均值与样本方差的重要定理。

定理 1.1 设 $X_1, \cdots, X_n$ 是来自总体 $X\sim N(\mu,\sigma^2)$ 的样本。μ,σ^2 均未知。记 $\overline{X}=\sum_{i=1}^n X_i/n$，$S^2=\sum_{i=1}^n (X_i-\overline{X})^2\big/(n-1)$，$U=(\overline{X}-\mu)\sqrt{n}/\sigma$。则：

（1）$\overline{X}\sim N(\mu,\sigma^2/n)$，$U=(\overline{X}-\mu)\sqrt{n}/\sigma\sim N(0,1^2)$；

（2）$(n-1)S^2\big/\sigma^2=\sum_{i=1}^n (X_i-\overline{X})^2\big/\sigma^2\sim\chi^2(n-1)$；

（3）$(\overline{X}-\mu)$与S^2相互独立;；

（4）$\sqrt{n}(\overline{X}-\mu)\big/S\sim t(n-1)$。

证明：（1）$\because E\overline{X}=\mu$，且$D\overline{X}=\sigma^2/n$，故$\overline{X}\sim N(\mu,\sigma^2/n)$，且$U\sim N(0,1^2)$。下证（2）与（3）。令$\mathbf{X}=(X_i,1\le i\le n)^T,\mathbf{Y}=(Y_i,1\le i\le n)^T$。设 A 为 n 阶正交方阵，其中第一行元素均为$1/\sqrt{n}$，及$\mathbf{Y}=A\mathbf{X}$。因为 A 为正交变换，故其平方和保持不变，即 $\sum_{i=1}^n X_i^2=\sum_{i=1}^n Y_i^2$，又$\det A=1$，而$\mathbf{X}=(X_i,1\le i\le n)^T$的 $p.d.f.$为：

$$(\sqrt{2\pi}\sigma)^{-n}\exp\{-\sum_{i=1}^{n}(x_i-\mu)^2/2\sigma^2\}=(\sqrt{2\pi}\sigma)^{-n}\exp\{-(\sum_{i=1}^{n}x_i^2-2\mu\sum_{i=1}^{n}x_i+n\mu^2)/2\sigma^2\}$$

注意到 A 的第一行元素均为 $1/\sqrt{n}$，因而 $y_1=\sum_{i=1}^{n}x_i/\sqrt{n}=\sqrt{n}\bar{x}$，其中 $\bar{x}=\sum_{i=1}^{n}x_i/n$，故由第三章随机向量变换前后 $p.d.f.$的关系式知 $\mathbf{Y}=(Y_i,1\le i\le n)^T$ 的 $p.d.f.$为：

$$\begin{aligned}&(\sqrt{2\pi}\sigma)^{-n}\exp\{-(\sum_{i=1}^{n}y_i^2-2\mu\sqrt{n}y_1+n\mu^2)/2\sigma^2\}\\&=(\sqrt{2\pi}\sigma)^{-1}\exp[-(y_1-\sqrt{n}\mu)^2/2\sigma^2]\prod_{i=2}^{n}[(\sqrt{2\pi}\sigma)^{-1}\exp(-y_i^2/2\sigma^2)]\end{aligned}$$

易知 $(Y_1,\cdots,Y_n)$ 为正态分布，且 $p.d.f.$的表达式是由 n 个 $p.d.f.$相乘而来，故 $Y_1,\cdots,Y_n$ 独立，且 $Y_1\sim N(\sqrt{n}\mu,\sigma^2),Y_i\sim N(0,\sigma^2),i=2,\cdots,n$。再由独立 $r.v.$的性质知 Y_1 与 $\sum_{i=2}^{n}Y_i^2$ 相互独立。而 $\sum_{i=2}^{n}Y_i^2=\sum_{i=1}^{n}Y_i^2-Y_1^2=\sum_{i=1}^{n}X_i^2-(\sum_{i=1}^{n}X_i)^2/n=\sum_{i=1}^{n}(X_i-\bar{X})^2=(n-1)S^2$，从而（3）得证。又注意到 $Y_i/\sigma\sim N(0,1^2),i=2,\cdots,n$，故 $(n-1)S^2/\sigma^2=\sum_{i=2}^{n}(Y_i/\sigma)^2\sim\chi^2(n-1)$

即(2)得证。下证(4)。由(3)知 $U=(\bar{X}-\mu)\sqrt{n}/\sigma$ 与 $(n-1)S^2/\sigma^2$ 独立，又由 $U\sim N(0,1^2)$，$(n-1)S^2/\sigma^2\sim\chi^2(n-1)$，故 $U/[((n-1)S^2/\sigma^2)/(n-1)]^{1/2}\sim t(n-1)$，即 $\sqrt{n}(\bar{X}-\mu)/S\sim t(n-1)$。证毕。

§2 点估计：极大似然估计与贝叶斯估计

统计估计是数理统计基本的研究内容之一。统计估计包括参数估计与非参数估计。在参数估计中又分为点估计与区间估计。本节给出有关点估计的一些重要概念与三种基本方法。

这里指的参数是指总体 X 的分布中的未知参数。所得参数估计是指用其样本的统计量对总体的（未知）参数作出估计。例如 X 为正态总体，其分布 $N(\mu,\sigma^2)$ 中的 μ,σ^2 未

知，则可用其样本 X_1, …, X_n 的统计量 $\overline{X}=\frac{1}{n}\sum_{i=1}^{n}X_i$ 及 $S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 分别作为未知参数 μ,σ^2 的估计。由于未知参数是 n 维空间的一个点，统计量的取值也是一个点，用后者的点去估计未知参数的点，这种估计称之为参数点估计。

一般地，设总体 X 的分布 $f(x,\theta)$，其中 $\theta=(\theta_i,1\le i\le n)$ 为未知参数，它的可能取值全体记为 Θ，称 Θ 为参数空间。记 X 的样本为 X_1, …, X_n，若取样本的统计量 $g(X_1, \cdots, X_n)$ 作为参数 θ 的估计，记作 $\widetilde{\theta}=g(X_1,\cdots,X_n)$，称 $g(X_1, \cdots, X_n)$ 为 θ 的估计量。若样本的观测值为 $x_1,\cdots,x_n$，称 $\widetilde{\theta}=g(x_1,\cdots,x_n)$ 为 θ 的估计值。

参数点估计的方法具体有很多，本节仅介绍最基本的三种。

注意：本书在数理统计部分，仅限于讨论离散型与连续型的两种总体。总体 X 的分布用 $f(x,\theta)$ 表示，如总体 X 为离散型 *r.v.*，则 $f(x,\theta)$ 表示分布律；如总体 X 为连续型 *r.v.*，则 $f(x,\theta)$ 表示概率密度函数，简称 $f(x,\theta)$ 为总体分布。

1 极大似然估计（*MLE*）

极大似然估计在历史上最早由高斯(*C.F.Gauss*)在研究误差理论中提出，到 1912 年由著名的统计学家费歇尔（*R.A.Fisher*）在一篇论文中把它作为一般参数估计方法提出来。随后他又与其它许多统计学家对该方法不断探索，至今已成为统计估计中最重要的方法。为了了解极大似然估计的直观依据与思想，先举几个例子：

例 2.1 甲，乙两人射击，已知甲击中目标的概率为 $p_1=0.1$，乙击中的概率为 $p_2=0.95$，假定在射击时，观察者只能看到射击结果是否击中目标，而且只知道有甲，乙两射手，但每次究竟是谁射击无法看到。现射击一枪，结果击中，试推断或估计这一枪来自哪位射手。

显然，若射击结果击中，则观察者估计这一枪来自乙射手较为合理，因为这一结果更像是乙似然。那么如何将这种直观的合理的推断提炼为一个估计的原理呢？

可以将上述由实验结果作出直观估计（猜测）的依据用数理统计的语言描述如下：令总体 $X\sim B(1,p)$ 服从二点分布，以 $(X=1)$ 表击中，$(X=0)$ 表未击中，及相应的概率为：$P(X=1)=p$，$P(X=0)=1-p$，则 X 的分布律为：

$$L(X,p)=p^X(1-p)^{1-X}\quad (X=0,1)$$

其中 p 是未知参数，如何用抽样结果来估计 p 是多少呢？

当实验结果($X=i$)发生时，$L(i,p)$ 看作是 p 的函数（i=0,1）。于是 $L(i,p)$ 可理解为，当参数 p 不同时，事件($X=i$)发生的可能性就不同。现在，事件(X=1)发生，由 $L(1,p_2)=p_2>p_1=L(1,p_1)$ 表明是乙射手使(X=1)发生的可能性要比甲射手使(X=1)发生的可能性要大很多。因此，取 p 的估计 $\tilde{p}=p_2=0.95$ 较为合理。

更进一步将上例一般化。设总体 $X\sim B(1,p)$，未知参数 $p\in[0,1]$。若取一个样本是事件($X=i$)发生（i=0,1），则未知参数 p 的估计 $\tilde{p}$ 应满足：

$$L(i,\tilde{p})=\max_{p\in[0,1]}L(i,p)\quad(i=0,1)$$

或：

$$L(X,\tilde{p})=\max_{p\in[0,1]}L(X,p)$$

若取容量是 n 的样本 X_1, …，X_n，其分布为：

$$L(X_1,\cdots,X_n;p)=\prod_{k=1}^{n}\left[p^{X_k}(1-p)^{1-X_k}\right]=p^{\sum\limits_{k=1}^{n}X_k}(1-p)^{n-\sum\limits_{k=1}^{n}X_k}\ .$$

如由试验结果 $X_1=i_1,\cdots,X_n=i_n$，来估计未知参数 p，则应取 $\tilde{p}$ 满足：

$$L(i_1,\cdots,i_n;\tilde{p})=\max_{p\in[0,1]}L(i_1,\cdots,i_n;p)$$

或：

$$L(X_1,\cdots,X_n;\tilde{p})=\max_{p\in[0,1]}L(X_1,\cdots,X_n;p)$$

这种使已发生的抽样结果达到最大的参数 $\tilde{p}$ 作为对 p 的估计，就是极大似然估计原理。

例 2.2 一个袋子里有 5 个球，既有红球也有白球，已知它们的数目之比是 $4:1$，但不知是红球多还是白球多。即随机抽取一个红球的概率是 $p_1=1/5$ 或者 $p_2=4/5$。现在有

放回地从袋子里抽取 3 个球，试用抽取的红球数来估计红球的比例是 p_1 还是 p_2。

解：记抽取的红球数为 X，易知 $X \sim B(3,p)$。当 $X=k$ $(0 \le k \le 3)$时，记它的概率为：

$$L(k,p) = C_3^k p^k (1-p)^{3-k}$$

它是 p 的函数，下表给出了 $p= p_1=1/5$ 与 $p= p_2=4/5$ 时 $P(X=k)$的值：

X	0	1	2	3
$p=1/5$ 时 $P(X=k)$ 的值	64/125	48/125	12/125	1/125
$p=4/5$ 时 $P(X=k)$ 的值	1/125	12/125	48/125	64/125

可知：当 k=0,1 时，$L(k,p_1) > L(k,p_2)$，故应取 $\tilde{p} = p_1$=1/5；当 k=2,3 时，$L(k,p_1) < L(k,p_2)$，故应取 $\tilde{p} = p_2$=4/5。即说明应取使得事件($X=k$)发生的可能性最大的参数作为我们的估计较为合理，符合客观推理，即应取 $\tilde{p}$ 满足：

$$L(k,\tilde{p}) = \max_p L(k,p) \quad 0 \le k \le 3$$

或：

$$L(X,\tilde{p}) = \max_p L(X,p)$$

例 2.3 已知一批零件的直径 $X \sim N(\mu,1^2)$，其中σ=1 已知。其概率密度是 $f(x;\mu)=\frac{1}{\sqrt{2\pi}} exp(-(x-\mu)^2/2)$，其中 μ 未知，但设 μ 有两种取值：μ_1= 1 或 μ_2=5，现抽取一个样本，若抽取的结果为 1.1，则估计总体的参数 μ 是 1 还是 5？

解：记 $L(x,\mu) = f(x,\mu)$，当给定 x 时，$L(x,\mu)$ 看作是 μ 的函数。现在抽样结果是 $x = 1.1$，在此条件下，比较两个参数 $\mu=\mu_1$= 1 与 $\mu=\mu_2$=5，则不难得出 $L(1.1,1) > L(1.1,5)$。这表明在 $x = 1.1$ 的前提下，未知参数 μ 取 $\mu = 1$ 的可能性要比取 μ=5 的可能性大，故选取 $\tilde{\mu} = 1$ 作为 μ 的估计较为合理。

以上三例估计参数的直观依据是：选取参数 p（或 μ）的估计 $\tilde{p}$（或 $\tilde{\mu}$）满足：使已观测到的样本出现的可能性最大的参数作为未知参数 p（或 μ）的估计，这种估计方法称为极大似然估计（*Maximum Likelihood Estimate*）简记为 *MLE*。一般情形叙述如下：

设总体的分布为 $f(x,\theta)$，未知参数 $\theta=(\theta_1,\cdots,\theta_m)\in\Theta$，其样本 X_1, ⋯, X_n 的分布为：$L(x;\theta)=L(x_1,\cdots,x_n;\theta)=\prod_{k=1}^{n}f(x_k,\theta)$，其中 $x=(x_1,\cdots,x_n)$，$\theta=(\theta_1,\cdots,\theta_m)\in\Theta$。显然，若 $L(X;\theta)>L(X';\theta)$，意味着给定θ，出现 X 要比出现 X' 的可能性大。若已观测到样本 $x_1,x_2,\cdots,x_n$，且对不同的参数 θ' 与 θ'' 有 $L(x_1,\cdots,x_n;\theta')>L(x_1,\cdots,x_n;\theta'')$，则意味着未知参数θ取 θ' 的可能性要比θ取 θ'' 的可能性要大。因此，用样本 $(X_1,\cdots,X_n)$ 来估计未知参数θ时，应取 $\tilde{\theta}^*$ 满足：

$$L(X_1,\cdots,X_n;\tilde{\theta}^*)=\max_{\theta\in\Theta}L(X_1,\cdots,X_n;\theta) \tag{2.1}$$

称 $\tilde{\theta}^*$ 为θ的极大似然估计，称θ的函数 $L(x_1,\cdots,x_n;\theta)$ 为似然函数。

显然，满足(2.1)式的 $\tilde{\theta}^*$ 是 $(X_1,\cdots,X_n)$ 的函数，记为 $\tilde{\theta}^*=\tilde{\theta}^*(X_1,\cdots,X_n)$，称 $\tilde{\theta}^*(X_1,\cdots,X_n)$ 为θ的极大似然估计量，称 $\tilde{\theta}^*(x_1,\cdots,x_n)$ 为θ的极大似然估计值。

注：若要估计的是未知参数θ的函数 $g(\theta)$，则取 $g(\tilde{\theta}^*)$为 $g(\theta)$的极大似然估计。

如何求极大似然估计呢？下面先介绍似然函数 $L(x;\theta)$ 关于θ有连续的一阶（偏）导数的情形。

此时，由于 $L(x;\theta)$ 与 $\ln L(x;\theta)$ 有相同的极值点，因此，若 $\tilde{\theta}^*$ 是θ的 *MLE*，则它应满足方程组：

$$\frac{\partial\ln L(x;\theta)}{\partial\theta_i}=0,\quad(1\le i\le m) \tag{2.2}$$

求解方程组(2.2)，且与边界值比较，其中达到 $L(x;\theta)$ 最大值对应的θ就是 $\tilde{\theta}^*$。称(2.2)为似然方程。

例 2.4 设样本 $X_1,\cdots,X_n$ 取自总体 $X\sim N(\mu,1^2)$，其中σ=1 已知，μ为未知参数。试求μ的 *MLE* $\tilde{\mu}^*$。

解：$X_1,\cdots,X_n$ 的似然函数为：

$$L(X_1,\cdots,X_n;\mu)=\prod_{k=1}^{n}\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(X_k-\mu)^2}$$

取对数，得

$$\ln L(X_1,\cdots,X_n;\mu)=-\frac{n}{2}\ln(2\pi)-\frac{1}{2}\sum_{i=1}^{n}(X_i-\mu)^2$$

对 $\ln L$ 关于μ求导，并令它为 0，得：$\frac{d\ln L}{d\mu}=\sum_{k=1}^{n}(X_k-\mu)=0$。不难验证，当 $\mu<\overline{X}=\sum_{k=1}^{n}X_k/n$ 时，$\frac{d\ln L}{d\mu}>0$；当 $\mu>\overline{X}=\sum_{k=1}^{n}X_k/n$ 时，$\frac{d\ln L}{d\mu}<0$。故 $\mu=\overline{X}$ 为 $L(X_1,\cdots,X_n;\mu)$ 的极大点，亦是最大值点。故 $\tilde{\mu}^*=\overline{X}$ 是μ的 *MLE*。

例 2.5 （均匀分布）设样本 $X_1,\cdots,X_n$ 取自总体 $X\sim U(0,\theta)$，（θ>0 为未知参数）。则似然函数：

$$L(X_1,\cdots,X_n;\mu)=\theta^{-n}\prod_{k=1}^{n}I_{(0<X_k<\theta)}$$

此时无法利用(2.2)求 *MLE*。但注意到由于 $\theta>X_k,\forall 1\le k\le n$，为使 L 达最大，当且仅当取：$\tilde{\theta}^*=\max_{1\le k\le n}X_k$。

以下是总体为离散型 *r.v.*的例子：

例 2.6 设样本 $X_1,\cdots,X_n$ 取自总体 $X\sim Po(\lambda)$，其中λ>0 为未知参数。试求其 $MLE\ \widetilde{\lambda}^*$。

解：X 的分布律为：$P(X=i)=(\lambda^i/i!)e^{-\lambda}$，$i=0,1,\cdots$，故似然函数为：$L(X_1,\cdots,X_n;\lambda)=\prod_{k=1}^{n}(\lambda^{X_k}/X_k!)e^{-\lambda}$，由 $\frac{d\ln L}{d\lambda}=0$，可得 $\widetilde{\lambda}^*=\overline{X}$

下面对几类常见的分布，找出它们的参数的极大似然估计。

（1）指数分布：$X\sim E_X(\lambda)$。$f(x)=\lambda e^{-\lambda x}I_{(x>0)}$，参数 $\lambda>0$。当样本取值 $(x_1,\cdots,x_n)$ 时，它的似然函数为：$L(x_1,\cdots,x_n;\lambda)=\lambda^n\prod_{i=1}^{n}e^{-\lambda x_i}=\lambda^n e^{-\lambda\sum_{i=1}^{n}x_i}$ 对数似然函数记为：$lnL=nln\lambda-\lambda\sum_{i=1}^{n}x_i$，$\frac{\partial\ln L_n}{\partial\lambda}=\frac{n}{\lambda}-\sum_{i=1}^{n}x_i$，故 $\widetilde{\lambda}^*=1/\overline{X}$ 就是λ的极大似然估计。

（2）正态分布：分布密度 $f(x;\ \mu,\ \delta)=\frac{1}{\sqrt{2\pi\delta}}e^{-\frac{1}{2\delta}(x-\mu)^2}$（其中 $\delta=\sigma^2>0$），样本的似然函数：$L(x_1,\cdots,x_n;\mu,\delta)=\left(\frac{1}{\sqrt{2\pi\delta}}\right)^n\prod_{i=1}^{n}e^{-\frac{1}{2\delta}(x_i-\mu)^2}=(2\pi)^{-n/2}\delta^{n/2}e^{-\frac{1}{2\delta}\sum_{i=1}^{n}(x_i-\mu)^2}$，故 $\ln L=-\frac{n}{2}\ln(2\pi)-\frac{n}{2}\ln\delta-\frac{1}{2\delta}\sum_{i=1}^{n}(x_i-\mu)^2$，则似然方程组为：

$$\begin{cases}\dfrac{\partial\ln L_n}{\partial\mu}=\dfrac{1}{\delta}\sum_{i=1}^{n}(x_i-\mu)=0\\ \dfrac{\partial\ln L_n}{\partial\delta}=-\dfrac{n}{2\delta}+\dfrac{1}{2\delta^2}\sum_{i=1}^{n}(x_i-\mu)^2=0\end{cases}$$

解得 $\mu=\frac{1}{n}\sum_{i=1}^{n}x_i=\bar{x}$，$\delta=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$。故 $\widetilde{\mu}^*=\overline{X}$，$\widetilde{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2$ 分别为 μ与σ^2 的极大似然估计量。

2 矩估计

考察常用的总体分布中，许多未知参数与它们的矩有密切的关系。例如正态总体 $X\sim N(\mu,\sigma^2)$ 中的参数 μ,σ^2 本身就是 X 的一阶原点矩和二阶中心矩。又如 $X\sim Po(\lambda)$，则

$\lambda=EX$。又如 $X\sim E_X(\lambda)$，此时 $EX=\lambda^{-1}$，可知参数λ是一阶矩的函数。由此，一种很自然的估计方法是用样本的矩作为总体相应矩的估计，再利用未知参数与总体矩的函数关系，可得未知参数的估计量，称这种方法为矩估计。

设 $X_1,\cdots,X_n$ 是总体 $X\sim F(x,\theta)$ 的样本，$\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ 为样本均值，$S^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_k-\overline{X})^2$ 为样本方差，$\frac{1}{n}\sum_{i=1}^{n}X_i^k$ 为样本 k 阶原点矩，$\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^k$ 为样本 k 阶中心矩。

例 2.7 设样本 $X_1,\cdots,X_n$，来自总体 $X\sim N(\mu,\sigma^2)$，其中 μ,σ^2 为未知参数。试求 μ,σ^2 的矩估计。

解：由样本矩和总体矩的定义可知，$\mu=EX$，故 $\tilde{\mu}=\frac{1}{n}\sum_{k=1}^{n}X_k=\overline{X}$ 为 μ 的矩估计。同理，$\tilde{\sigma}^2=\frac{1}{n}\sum_{i=1}^{n}(X_k-\overline{X})^2=S_1^2$ 为 σ^2 的矩估计。又由 $\sigma=\sqrt{DX}$ 可知，$\tilde{\sigma}=S_1$ 是σ的矩估计。

例 2.8 设总体 $X\sim B(n,p)$，其中 n 已知，求 p 的矩估计。

解：由 $EX=np$，可得：$\overline{X}=\tilde{p}n$，故 $\tilde{p}=\overline{X}/n$ 是 p 的矩估计。

例 2.9 设总体 X 是Γ分布（参看第二章），即 *p.d.f.*为 $f(x)=\frac{1}{\Gamma(\alpha)}\beta^{\alpha}x^{\alpha-1}e^{-\beta x}I_{(x\ge 0)}$，其中 α,β 为未知参数。试求 α,β 的矩估计量。

解：经计算可知：$EX=\alpha/\beta, DX=\alpha/\beta^2$，故应取 $\overline{X}=\tilde{\alpha}/\tilde{\beta}, S^2=\tilde{\alpha}/\tilde{\beta}^2$，解之得：$\tilde{\alpha}=\overline{X}^2/S^2,\tilde{\beta}=\overline{X}/S^2$。

3 贝叶斯（*Bayes*）估计

本小节简要讨论一下贝叶斯（*T.R.Bayes*）估计的基本思想。

设总体的分布为 $f(x,\theta)$，其中 $\theta\in\Theta$ 为未知参数，问题是用样本 X_1, ⋯, X_n 对参数 θ 作出估计。以前我们所讨论的极大似然估计，矩估计等方法中，假设未知参数θ只是简单

的一个未知数，在抽取样本之前，我们对θ没有任何了解，所有的信息完全来自样本。

而贝叶斯估计法的出发点是：首先，把未知参数θ看作是随机变量（或随机向量），且在抽样之前，对θ已有一定的知识（包括合理的猜想与准则），称其为先验知识。这种先验知识用θ的某种概率分布表达出来，记为 $h(\theta)$。这种概率分布叫做 θ 的“先验分布”或“验前分布”。这个分布反映了在试验之前对未知参数 θ 的所把握的信息。其次，将原来含有参数θ的样本 $X_1,\cdots,X_n$ 分布 $f(x_1,\theta)\cdots f(x_n,\theta)$ 定义为在给定 θ 值时，（ $X_1,\cdots,X_n$ ）的条件分布。因而（ $\theta,X_1,\cdots,X_n$ ）的联合密度为：$h(\theta)\ f(X_1,\theta)\cdots f(X_n,\theta)$，于是，$(X_1,\cdots,X_n)$ 的边缘密度为：

$$p(X_1,\cdots,X_n)=\int_{\theta\in\Theta}h(\theta)f(X_1,\theta)\cdots f(X_n,\theta)d\theta$$

第三，由此，*Bayes* 提出在给定 $X_1,\cdots,X_n$ 的条件下，θ 的条件密度为

$$h(\theta|X_1,\cdots,X_n)=h(\theta)f(X_1,\theta)\cdots f(X_n,\theta)/p(X_1,\cdots,X_n) \tag{2.3}$$

称（2.3）式为 θ 的“后验密度”。这个条件密度代表了我们现在（即在取得样本 $X_1,\cdots,X_n$ 后）对 θ 的知识，它综合了 θ 的先验信息（ $h(\theta)$ 所反映的信息）与由样本带来的信息。读者不难看出（2.3）式在形式上与第一章的 *Bayes* 公式相类似。

最后利用后验分布 $h(\theta|X_1,\cdots,X_n)$ 作出对θ的推断（估计θ或对θ作检验）。

贝叶斯学派认为：在有了后验分布(2.3)后，对参数的估计，必须建立在这个后验分布的基础上推出，具体可结合使用者的某种要求（准则）处理。此处给出较常用的两种方法：（1）若取后验分布(2.3)的（条件）均值作为 θ 的估计，则称此为 *Bayes* 条件期望估计；（2）若取 $\tilde{\theta}_0$ 满足：后验分布(2.3)达最大的 θ 作为 $\tilde{\theta}_0$，称 $\tilde{\theta}_0$ 为 *Bayes* 最大后验估计。

例 2.10 某射手进行独立重复射击试验，设每次射击击中的概率为 p(未知)，现射击 n 次，击中 r 次。试给出 p 的 *Bayes* 条件期望估计。

解：由题意设总体 $X\sim B(1,p)$，记 X_i 表示第 i 次射击，(X_i=1)表示击中，(X_i=0)表示没有击中 (i=1, ⋯ ,n), $P(X_i=1)=p$，$P(X_i=0)=1-p$。因此 $X_1,\cdots,X_n$ 的分布律为

$p^x(1-p)^{n-x} \quad (x=0,1)$，其中 $X=\sum_{i=1}^{n}X_i=x$ 表击中次数。取 p 的先验密度 $h(p)$,则 p 的后验密度为

$$h(p|X_1,\cdots,X_n)=h(p)p^{\sum_{i=1}^{n}X_i}(1-p)^{n-\sum_{i=1}^{n}X_i}\Bigg/\int_0^1 h(p)p^{\sum_{i=1}^{n}X_i}(1-p)^{n-\sum_{i=1}^{n}X_i}dp \quad 0\le p\le 1 \qquad (2.4)$$

若对该射手射击水平很不了解，可假设他的命中概率 p 在[0,1]上取每个值是等可能的，此时 $h(p)=I_{[0,1]}$ 是[0,1]上的均匀分布。代入(2.4)式有：

$$h(p|X_1,\cdots,X_n)=p^X(1-p)^{n-X}\Bigg/\int_0^1 p^X(1-p)^{n-X}dp$$

即 $h(p|X_1,\cdots,X_n)$ 为贝塔分布。如用后验分布的期望值，即 p 对 $X_1,\cdots,X_n$ 的条件期望 $E(p|X_1,\cdots,X_n)$作为 p 的估计量，即得：

$$\tilde{p}=\tilde{p}(X_1,\cdots,X_n)=\int_0^1 ph(p|X_1,\cdots,X_n)dp=(X+1)/(n+2) \qquad (2.5)$$

$\tilde{p}$ 即为 p 的贝叶斯估计。

现比较上例 p 的三个不同估计。n 次射击，命中 X 次。用极大似然估计和矩估计都是用 n 次射击命中的频率 $X/n=\tilde{p}$ 去估计 p。这种估计有它的不足之处，例如，当 $n=X=1$ 时，估计 $\tilde{p}=1$；而当 $n=X=100$ 时，此估计还是 $\tilde{p}=1$。再有一种极端情形，对 $n=1$，而 $X=0$，得 $\tilde{p}=0$；对 $n=100$，$X=0$，仍是 $\tilde{p}=0$。射击 100 次每次都命中，直觉上此射手命中概率相当大，此时估计 $\tilde{p}=1$ 合情合理；倘若仅射一次，命中了，但此时该射手命中概率不应与射击 100 次每次命中相提并论，遗憾的是 $\tilde{p}=X/n$ 作为 p 的估计其结果都是一样的。然而，用贝叶斯估计量 p 的估计是 $\tilde{p}=(X+1)/(n+2)$。当 $n=X=1$ 时，$\tilde{p}=2/3$；而当 $n=X=100$ 时，$\tilde{p}=101/102$。显然，这个估计比用 X/n 要合理，更符合多数人的直观估计。

例 2.11 设总体 $X\sim N(\mu,\sigma^2)$，其中 σ 已知，而μ为未知参数，μ的先验分布为 $N(\mu_0,\sigma_0^2)$，其中 $\mu_0,\sigma_0^2>0$ 为已知常数，$X_1,\cdots,X_n$ 为 X 的样本，试求μ的 *Bayes* 条件期望估计。

解：此时 $h(\mu)=(\sqrt{2\pi}\sigma_0)^{-1}\exp(-(\mu-\mu_0)^2/2\sigma_0^2)$，

$$f(x,\mu)=(\sqrt{2\pi}\sigma)^{-1}\exp(-(x-\mu)^2/2\sigma^2)$$

由(2.3)式知，μ的后验密度（即μ关于 $X_1,\cdots,X_n$ 的条件密度）经整理后为：

$$h(\mu|X_1,\cdots,X_n)=K\exp\{-\frac{1}{2}[(\mu-\mu_0)^2/\sigma_0^2+\sum_{k=1}^{n}(X_k-\mu)^2/\sigma^2]\}$$

其中 K 是一个与μ无关，只与 $\mu_0,\sigma_0,X_1,\cdots,X_n$ 有关的数。记 $\overline{X}=\sum_{k=1}^{n}X_k/n$，将上式右端方括号的式子化简为：

$$\sum_{k=1}^{n}(X_k-\mu)^2/\sigma^2+(\mu-\mu_0)^2/\sigma_0^2=(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})[\mu-(n\overline{X}/\sigma^2+\mu_0/\sigma_0^2)(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})^{-1}]^2+c$$

其中 c 是一个与μ无关，只与 $\sigma,\mu_0,\sigma_0,X_1,\cdots,X_n$ 及 n 有关的数。故：

$$h(\mu|X_1,\cdots,X_n)=K\exp\{-\frac{1}{2}(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})[\mu-(n\overline{X}/\sigma^2+\mu_0/\sigma_0^2)(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})^{-1}]^2\}+c \tag{2.6}$$

注意上式说明 $h(\mu|X_1,\cdots,X_n)$ 是正态分布，且条件期望为：$(n\overline{X}/\sigma^2+\mu_0/\sigma_0^2)(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})^{-1}$，故得μ的 *Bayes* 的条件期望估计为：

$$\tilde{\mu}=(n\overline{X}/\sigma^2+\mu_0/\sigma_0^2)(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})^{-1}$$

即：

$$\tilde{\mu}=\left(\frac{n}{n+(\sigma^2/\sigma_0^2)}\right)\overline{X}+\left(\frac{(\sigma/\sigma_0)^2}{n+(\sigma^2/\sigma_0^2)}\right)\mu_0 \tag{2.7}$$

将μ的 *Bayes* 条件期望估计 $\tilde{\mu}$ 分解成(2.7)颇有意味。考虑两种极端情形下的估计：若仅有实验样本结果的信息，而毫无先验信息，则此同于以往矩估计与极大似然估计用样本均值去估计μ。另一种是若仅有先验信息 $h(\mu)\sim N(\mu_0,\sigma_0^2)$，而无取样本的信息，这是只好用先验分布的均值$\mu_0$作为μ的估计。而当同时掌握了两种信息，则 *Bayes* 的条件期望估计是上述两种极端情形的“加权” 平均，“权”的比值为 $n:(\sigma/\sigma_0)^2$。当 $\sigma=\sigma_0$ 时，权之比为 $n:1$，即如果先验知识中μ的方差与总体的方差一样时，则样本信息与先验信息对μ的估计占的比例为 $n:1$。当 $\sigma/\sigma_0=\sqrt{n}$ 时，权之比为 $1:1$，此时样本信息与先验信息对μ的估计而言占同等重要地位！足见 *Bayes* 估计的合理性！

再从上面例 2.11 来看对同一参数的 *Bayes* 估计与极大似然估计的关系：注意 σ_0^2 是以往“经验值” μ_0 的方差，它是刻划 μ_0 的“精度”的。若 $\sigma_0^2\to\infty$，则意味着对μ而言，已没有任何先前经验可言。此时，由（2.7）式，当 $\sigma_0^2\to\infty$ 时，μ的 *Bayes* 估计 $\tilde{\mu}$ 就等于μ的极大似然估计！这说明，对这一问题而言，前者的结果包含了后者。

另外，还要强调一点的是：*Bayes* 统计首先把被估计参数 θ 看作是随机变量是否恰当，是否合理？在许多具体问题中，把参数 θ 看作是 *r.v.* 更加符合客观实际。例如，大批量生产的某一元件，设次品率为 p，将这批元件打包成箱销售，每箱 1000 个。显而易见，此时各箱的次品率 θ 就是一个随机变量。同时，在理论上，假设参数 θ 是一个随机变量，这个观点本身并无不妥。当然，与任何其它数理统计一样，*Bayes* 统计理论还有待进一步完善，改进与发展。从应用的角度看，在没有先验资料与数据的条件下，如何恰当选取其先验分布，这是关键。因此，值得进一步深入探讨。

§3 估计的优良性准则

同一未知参数的估计量有许多不同的方法，那么怎样来衡量与比较估计量的好坏呢？又什么样的估计量是较好的呢？这就涉及估计量的优良性标准。粗略的说，估计量与被估计量越“接近”越好，即“误差”越小越好。

设 X 的密度函数为 $f(x,\theta)$，其中 $\theta=(\theta_1,\cdots,\theta_m)\in\Theta$，$\Theta$ 是 R^m 的非空集合。设 $g(\theta)$ 是 θ 的函数，$X_1,\cdots,X_n$ 是 X 的样本。所谓 $g(\theta)$ 的估计量，是指样本函数 $\varphi(X_1,\cdots,X_n)$ 的不同选择可得到不同的估计量。直观上看，$|\varphi(X_1,\cdots,X_n)-g(\theta)|$ 越小，φ 就越好。但

$\varphi(X_1,\cdots,X_n)$ 的值是依赖于样本值的，它本身是随机变量，而 $g(\theta)$ 是未知的。所以评价估计量的优劣需要衡量优良性的标准。

为简化记号，以下只讨论对参数 $\theta=(\theta_1,\cdots,\theta_m)$ 的估计问题，这时因为对 θ 的函数 $\theta'=g(\theta)$ 的估计问题完全可化为对参数 θ' 的估计，且简记 $\widetilde{\theta}=\varphi(X_1,\cdots,X_n)$ 取为 θ 的估计量。我们从估计量的几个不同侧面的概率特性及估计量所产生的效果来刻划估计量的优劣。

1 无偏估计

定义 3.1 称 $\widetilde{\theta}=\varphi(X_1,\cdots,X_n)$ 是参数 θ 的无偏估计（*Unbiased Estimate*），若：

$$E\widetilde{\theta}=\theta \qquad (\theta\in\Theta)$$

对估计量的无偏性要求在通常情形下是合理的要求，但对某些情形下不见得是必要的。

例 3.1 设 $(X_1,\cdots,X_n)$ 为总体 X 的样本，记μ=*EX*，若 $DX=\sigma^2$ 存在。则样本均值 $\overline{X}=\sum\limits_{i=1}^{n}X_i/n$ 是μ的无偏估计，样本方差 $S^2=\sum\limits_{i=1}^{n}(X_i-\overline{X})^2/(n-1)$ 是 σ^2 的无偏估计。

证明：因 $E\overline{X}=E(\sum\limits_{i=1}^{n}X_i)/n=\mu$，故 $\overline{X}$ 是μ的无偏估计，而

$$ES^2=E\left\{\sum_{i=1}^{n}(X_i-\overline{X})^2/(n-1)\right\}=E\left\{\sum_{i=1}^{n}[(X_i-\mu)-(\overline{X}-\mu)]^2\right\}/(n-1)$$

$$=E\left\{\sum_{i=1}^{n}(X_i-\mu)^2-n(\overline{X}-\mu)^2\right\}/(n-1)=\left\{\sum_{i=1}^{n}DX_i-nD\overline{X}\right\}/(n-1)=\sigma^2 \quad (\text{注}D\overline{X}=\sigma^2/n)$$

故 S^2 是 σ^2 的无偏估计。

无偏性只有在估计量重复使用时才有意义。例如，估计量 $\widetilde{\theta}$ 独立地重复利用 m 次 $\widetilde{\theta}^{(1)},\cdots,\widetilde{\theta}^{(m)}$ 作为 θ 的估计。若 $E\widetilde{\theta}=\theta$，则由大数定律知其平均值 $\frac{1}{m}\sum_{i=1}^{m}\widetilde{\theta}^{(i)}\xrightarrow{p}\theta \quad (m\to\infty)$（依概率收敛)。这意味着多次反复利用估计量 $\widetilde{\theta}$ 与真值 θ 没有系统偏差，因而无偏性刻划了多次使用估计量 $\widetilde{\theta}$ 的“平均”效果。

无偏性的要求有时不见得是必要的，见本章练习题 10.10。

2 有效估计与最小方差无偏估计

一个参数估计可以有许多不同的无偏估计量。例如在例 3.1 中 $X_1, \alpha X_2+(1-\alpha)X_3$（其中 $0\le\alpha\le1$），$\overline{X}$ 及 $\sum_{i=1}^{n}\alpha_i X_i$（其中 $\sum_{i=1}^{n}\alpha_i=1$）均是μ的无偏估计。在众多的μ的无偏估计中，如何选择较好的？较优的？

定义 3.2 设 $\widetilde{\theta}_1$ 与 $\widetilde{\theta}_2$ 均是 θ 的无偏估计，若对于任意样本容量都有：

$$D\widetilde{\theta}_1 < D\widetilde{\theta}_2$$

则称 $\widetilde{\theta}_1$ 比 $\widetilde{\theta}_2$ 有效。

定义 3.3 设 $\widetilde{\theta}_0$ 是 θ 的无偏估计，若对 θ 的任何一个无偏估计 $\widetilde{\theta}$ 都有：

$$D\widetilde{\theta}_0 < D\widetilde{\theta} \qquad (\forall\theta\in\Theta)$$

则称 $\widetilde{\theta}_0$ 是 θ 的最小方差无偏估计量（*Minimum Variance Unbiased Estimate*）简称为 *MVU* 估计。

用以上定义要验算一个估计量是否为最小方差无偏估计量是难于实现的。若要深入讨论最小方差无偏估计量的求法，须涉及数理统计其他许多内容，故本书不作介绍。下面仅给出它的下界，即著名的克拉美—罗（*H.Cramer—C.r.Rao*）不等式。

定理 3.1 设总体 X 是连续型 *r.v.*，其 *p.d.f* 为 $f(x,\theta),\theta\in\Theta$，其中 Θ 是 R 上的一个开区间，$X_1,\cdots,X_n$ 是 X 的简单样本。$\widetilde{\theta}=\widetilde{\theta}(X_1,\cdots,X_n)$ 是 θ 的无偏估计。记 $x=(x_1,\cdots,x_n)$，$L(x;\theta)=\prod_{i=1}^{n}f(x_i,\theta)$，

$$I(\theta)\triangleq\int_R[(\frac{\partial f(u,\theta)}{\partial\theta})^2/f(u,\theta)]du \tag{3.1}$$

若满足：

（1）集合 $S_\theta=\{x_i, f(x_i,\theta)\ne0\}$ 与 θ 无关；

（2）$\dfrac{\partial f(x_i,\theta)}{\partial\theta}$ 存在，且对 $\forall\theta\in\Theta$ 有：

$$\frac{\partial}{\partial\theta}\int_R f(x_i,\theta)dx_i=\int_R\frac{\partial f}{\partial\theta}dx_i\,,\quad \frac{\partial}{\partial\theta}\int_{R^n}\widetilde{\theta}(x)L(x;\theta)dx=\int_{R^n}\widetilde{\theta}(x)\frac{\partial L}{\partial\theta}dx$$

（3）$I(\theta)>0 \quad \forall\theta\in\Theta$

则：

$$D\tilde{\theta}\geq\frac{1}{nI(\theta)} \tag{3.2}$$

证明：从略。可参看[3]第二册。

不等式(3.2)右边称为 *Cramer—Rao* 下界。而

$$I(\theta)=E(\frac{\partial\ln f(X,\theta)}{\partial\theta})^2=\int_R[(\frac{\partial\ln f(u,\theta)}{\partial\theta})^2 f(u,\theta)]du=\int_R[(\frac{\partial f(u,\theta)}{\partial\theta})^2/f(u,\theta)]du$$

最早是由 *R.A.Fisher* 在 20 世纪 20 年代提出的，故称 $I(\theta)$ 为 *Fisher* 信息量。$I(\theta)$ 越大，θ 的无偏估计越有可能达到更精确些。而 $\tilde{\theta}=\tilde{\theta}(X_1,\cdots,X_n)$ 是样本 $X_1,\cdots,X_n$ 的函数。θ 能估的越准，表明其样本所含参数 θ 的信息量越大。当样本容量为 n 时，(3.2)式右边分母是 $nI(\theta)$，因此每个样本提供的信息量为 $I(\theta)$。

若总体 X 为离散型 *r.v.*，分布律为：$p(i;\theta)=P(X=i;\theta) \quad \theta\in\Theta$，则 $I(\theta)=E(\frac{\partial\ln p(X,\theta)}{\partial\theta})^2$，若满足定理 3.1 种类似条件，则(3.2)式仍成立。

定义 3.4 若 $\tilde{\theta}_0$ 是 θ 的无偏估计，且 $D\tilde{\theta}_0=\frac{1}{nI(\theta)}$，则称 $\tilde{\theta}_0$ 是 θ 的有效估计。

例 3.2 设 $X_1,\cdots,X_n$ 是来自总体 $X\sim N(\theta,\sigma^2)$ 的样本，θ,σ^2 未知。试问 $\overline{X}$ 与 S^2 是否分别是 θ 与 σ^2 的最小方差无偏估计？

解：须分别计算 θ 与 σ^2 的 *Cramer—Rao* 下界。本题中，

$$f(x,\theta,\sigma^2)=(\sqrt{2\pi}\sigma)^{-1}\exp(-(x-\theta)^2/2\sigma^2)$$

$$I(\theta)=E(\frac{\partial\ln f(X,\theta,\sigma^2)}{\partial\theta})^2=(\sqrt{2\pi}\sigma)^{-1}\int_{-\infty}^{+\infty}\frac{1}{\sigma^4}(x-\theta)^2\exp(-(x-\theta)^2/2\sigma^2)dx=\frac{1}{\sigma^2}$$

$E\overline{X}=\theta,\ \ D\overline{X}=\frac{\sigma^2}{n}=\frac{1}{nI(\theta)}$，故 $\overline{X}$ 是 θ 的 *MVU* 估计。对于 σ^2，$ES^2=\sigma^2$，S^2 是 σ^2 的无偏估计，又 $I(\sigma^2)=E(\frac{\partial\ln f(X,\theta,\sigma^2)}{\partial\sigma^2})^2=\frac{1}{2\sigma^4}$，但 $DS^2=\frac{2}{n-1}\sigma^4>\frac{2}{n}\sigma^4=\frac{1}{nI(\theta)}$，故 S^2 不是 σ^2 的最小方差无偏估计。

例 3.3 设总体 $X\sim E_X(\frac{1}{\theta})$，$\theta>0$ 未知。试问 $\overline{X}$ 是否是 θ 的 *MVU* 估计？

解：显然，由于 $E\overline{X}=EX=\theta$，知 $\overline{X}$ 是 θ 的无偏估计。又 $D\overline{X}=\frac{1}{n}DX=\frac{\theta^2}{n}$，而

$$I(\theta)=\int_0^{+\infty}(\frac{\partial\ln f(x,\theta)}{\partial\theta})^2 f(x,\theta)dx=\int_0^{+\infty}(-1/n+x/n^2)-\frac{1}{\theta}e^{-x/\theta}dx=1/\theta^2$$

$\Rightarrow D\overline{X}=\frac{1}{nI(\theta)}=\frac{\theta^2}{n}$，故 $\overline{X}$ 是 θ 的 *MVU* 估计。

例 3.4 设总体 $X\sim B(1,p)$，$p\in(0,1)$未知。试问 $\overline{X}$ 是否是 p 的 *MVU* 估计？

解：$E\overline{X}=p$，故 $\overline{X}$ 是 p 的无偏估计。又 X 的分布律为 $p^X(1-p)^{1-X}$，

$$I(p)=\left(\frac{\partial\ln(1-p)}{\partial p}\right)^2(1-p)+\left(\frac{\partial\ln p}{\partial p}\right)^2 p=1/p(1-p)，\quad D\overline{X}=p(1-p)/n=1/nI(p)，$$

故 $\overline{X}$ 是 p 的 *MVU* 估计。

3 相合估计

由于估计量 $\tilde{\theta}=\varphi(X_1,\cdots,X_n)$ 是样本的函数，因而它依赖于样本容量 n。自然有兴趣也有必要考虑 $n\to\infty$ 时估计量 $\tilde{\theta}$ 的性态。下面本小节为突出 $\tilde{\theta}$ 与 n 的关系，故记 $\tilde{\theta}_n \triangleq \varphi(X_1,\cdots,X_n)$。显然，当 $n\to\infty$ 时，$\tilde{\theta}_n$ 是否以概率收敛到θ 至关重要。

定义 3.5 设 $\tilde{\theta}_n=\varphi(X_1,\cdots,X_n)$ 是θ的一个估计量，若 $\lim\limits_{n\to\infty}\tilde{\theta}_n \overset{p}{=}\theta$，即 $\forall\varepsilon>0$，有 $\lim\limits_{n\to\infty}P(|\tilde{\theta}_n-\theta|<\varepsilon)=1$，则称 $\tilde{\theta}_n$ 是θ的相合估计量（*Consistant estimate*），或称一致估计量。

显然，对于任意分布的总体 X，若 $DX<\infty$，则 $\overline{X}$ 是 EX 的相合估计，S^2 是 DX 的相合估计。相合性对于大样本的估计是最重要的衡量标准之一。

在 *Bayes* 统计的估计中，更多的是以估计产生的“效果”（可以是性能方面的，也可以是经济上的）来衡量估计与决策的优良性准则，这就是损失函数与风险函数。

4.损失函数与风险函数

（1）损失函数

设 $\tilde{\theta}$ 为参数 $\theta\in\Theta$ 的一个估计量，称非负二元函数 $L(\theta,\tilde{\theta})$ 为损失函数。例如

$L(\theta,\widetilde{\theta})=(\theta-\widetilde{\theta})^2$，或 $L(\theta,\widetilde{\theta})=|\theta-\widetilde{\theta}|$。它们表示若用 $\widetilde{\theta}$ 去估计 θ 的效果越差，其损失就越大。注意由于 θ 与 $\widetilde{\theta}=\varphi(X_1,\cdots,X_n)$ 是 *r.v.*，故 $L(\theta,\widetilde{\theta})$ 也是 *r.v.*。

（2）风险函数

风险函数是评价估计的一个准则，风险小估计的效果越好。称 $R(\theta,\widetilde{\theta})=EL(\theta,\widetilde{\theta})$ 为估计量 $\widetilde{\theta}$ 的风险（*risk*）函数。

（3）后验风险函数

给定样本 $X_1,X_2,\cdots,X_n$，称 $r_\theta(X_1,\cdots,X_n)=E(L(\theta,\widetilde{\theta})\,|\,X_1,\cdots,X_n)$ 为估计量 $\widetilde{\theta}$ 的后验风险（*risk*）函数。后验风险是估计量 $\widetilde{\theta}$ 的条件平均损失。

由条件期望的性质易知：$R(\theta,\widetilde{\theta})=E(r_\theta(X_1,\cdots,X_n))$。这样为求风险 $R(\theta,\widetilde{\theta})$ 可先求在给定样本 $X_1,\cdots,X_n$ 下，$L(\theta,\widetilde{\theta})$ 对 $X_1,\cdots,X_n$ 的条件数学期望 $r_\theta(X_1,\cdots,X_n)$，其次对 $r_\theta(X_1,\cdots,X_n)$ 关于样本 $X_1,\cdots,X_n$ 求期望即得 $R(\theta,\widetilde{\theta})$。

在 *Bayes* 统计中，如何从后验分布出发进而给出该参数的估计？在上一节已给出了两种较常用的办法，其中一个是从直观定义出发构造参数的估计量，如通常可用来刻划后验分布的数字特征，如后验分布的条件数学期望（均值）或后验分布的中位数 $M(\theta\,|\,X_1,\cdots,X_n)$ 来估计 θ。（设 *r.v.X*，若存在 *M*，满足：$P(X>M)=P(X\le M)=\dfrac{1}{2}$，称 *M* 为 *X* 的中位数）。这里再给出另一办法，就是从估计产生的效果出发提出适当的准则。这就是最小风险估计。

定义 3.6 设 $\widetilde{\theta}$ 为参数 θ 的一个估计量，$L(\theta,\widetilde{\theta})$ 为损失，$R(\theta,\widetilde{\theta})=EL(\theta,\widetilde{\theta})$ 为估计风险。若 $\widetilde{\theta}_b$ 为 θ 的一个估计量，满足：对 θ 的任一估计 $\widetilde{\theta}$ 有：

$$R(\theta,\widetilde{\theta}_b)\le R(\theta,\widetilde{\theta})$$

则称 $\widetilde{\theta}_b$ 为 θ 的最小风险估计。我们通常把最小风险估计称为 *Bayes* 估计。

可以证明，后验风险最小估计等于风险最小估计。因此，上述定义的 *Bayes* 估计就是后验风险最小估计。若取 $L(\theta,\widetilde{\theta})=(\widetilde{\theta}-\theta)^2$，则 *Bayes* 估计 $\widetilde{\theta}_b=E(\theta|X_1,\cdots,X_n)$ 。

例 3.5 设 $X\sim B(1,p)$，其中参数 p 未知，样本 $X_1,X_2,\cdots,X_n$，取 p 的先验估计为[0,1]上的均匀分布，损失函数取 $L(\theta,\widetilde{\theta})=(\theta-\widetilde{\theta})^2$，求 θ 的 *Bayes* 估计。

解：由 θ 与 $X_1,\cdots,X_n$ 的联合 $p.d.f.$ 为 $p^{\sum_{i=1}^{n}x_i}(1-p)^{n-\sum_{i=1}^{n}x_i}$， $p\in[0,1]$。记 $X=\sum_{i=1}^{n}X_i$，

则后验分布为 $h(p|X_1,\cdots,X_n)=\dfrac{\Gamma(n+1)}{\Gamma(X+1)\Gamma(n-X+1)}p^X(1-p)^{n-X}$ 。

因而 $\hat{p}_b=E(p|X_1,\cdots,X_n)=(X+1)/(n+2)=(nX+1)/(n+2)$ 。

§4 区间估计

前面几节介绍的参数点估计使用一个点（即一个数）去估计未知参数，然而在应用中常常要对某未知参数估计它的范围，这就是区间估计，顾名思义，区间估计是用一个区间去估计未知参数，即把未知参数值估计在某两界限之间。有时参数的范围估计比点估计更有意义。

例如，某一地区的天气预报下周平均气温在 22°*C* 到 30°*C* 之间；洪水期间长江某检测站预报下周水位在 17 米~22 米之间。这种区间估计是人们生活中更需要了解的信息。

区间估计粗略地说是用两个估计量 $\widetilde{\theta}_1,\widetilde{\theta}_2(\widetilde{\theta}_1\le\widetilde{\theta}_2)$ 作为端点的区间 $[\widetilde{\theta}_1,\widetilde{\theta}_2]$ 作为参数 θ 取值范围的估计。显然，这种估计有两个基本要求：首先，“精度”要求，即 $\widetilde{\theta}_2-\widetilde{\theta}_1$ 要尽量小，太大没有意义；其次，这个估计要有尽可能高的可信程度（即可靠度）。因此 $\widetilde{\theta}_2-\widetilde{\theta}_1$ 不能太小。我们希望既能得到较高的精度，又能得到较高的可靠度。但在样本密度

n 给定时，这两方面要求相互矛盾，丁.奈曼在 20 世纪 30 年代提出至今仍广泛应用的原则是：首先保证可靠度，在此前提下使精度尽量高。

定义 4.1 对于参数 θ，如果有两个统计量 $\tilde{\theta}_1=\tilde{\theta}_1$（$X_1,X_2\cdots,X_n$），$\tilde{\theta}_2=\tilde{\theta}_2$（$X_1,X_2\cdots,X_n$）满足对给定的 $\alpha\in(0,1)$,有

$$P(\tilde{\theta}_1\le\theta\le\tilde{\theta}_2)=1-\alpha$$

则称 $[\tilde{\theta}_1,\tilde{\theta}_2]$ 是 θ 的一个区间估计（或置信区间），$\tilde{\theta}_1$，$\tilde{\theta}_2$ 分别称为置信下限，置信上限，$1-\alpha$ 称为置信水平。

这里的置信水平，就是对可信程度的度量。置信水平为 $1-\alpha$，如取 $1-\alpha=95\%$, 就是说若对某一参数 θ 取 100 个容量为 n 的样本，用相同方法做 100 个置信区间 $(\tilde{\theta}_1^{(k)},\tilde{\theta}_2^{(k)})$ k=1,2,⋯,100,则其中约有 95 个区间包含了真参数 θ。因此，当我们实际上只做一次区间估计时，有理由认为它包含了真参数。这样判断当然有可能犯错误，但犯错误的概率只有 5%。

1 正态总体均值的区间估计

（1）标准差 σ 已知的情形

设正态总体 $X\sim N(\mu,\sigma^2)$，σ 已知。从总体抽得简单样本 $X_1,\cdots,X_n$，试求置信水平为(1−α)的μ的置信区间。

由前面的讨论可知，可用 $\overline{X}$ 作为μ的估计，易知 $\overline{X}\sim N(\mu,\sigma^2/n)$，故 $U=(\overline{X}-\mu)\sqrt{n}/\sigma\sim N(0,1^2)$，当给定α(0<α<1)时，存在 $u_{\frac{\alpha}{2}}$（$u_{\frac{\alpha}{2}}$ 可用标准正态分布表查出）满足：$P\{|U|<u_{\frac{\alpha}{2}}\}=1-\alpha$，即 $P\{|(\overline{X}-\mu)\sqrt{n}/\sigma|<u_{\frac{\alpha}{2}}\}=1-\alpha$，得

$$P\{\overline{X}-u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\}=1-\alpha \tag{4.1}$$

因此可取：

$$(\overline{X}-u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}},\overline{X}+u_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}) \tag{4.2}$$

作为μ的(1−α)置信水平的置信区间。$\overline{X}-u_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}$与$\overline{X}+u_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}}$分别称为μ的置信上限于置信下限。

通常取α=0.01,0.05 或 0.10 等。(4.1)式表明，随机区间$(\overline{X}-u_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}},\overline{X}+u_{\frac{\alpha}{2}}\dfrac{\sigma}{\sqrt{n}})$套住（覆盖）参数μ的概率为(1−α)。

注意：给定α(0<α<1)，满足(4.1)式的置信区间不唯一。

（2）标准差σ未知的情形

由于σ未知，需选用$(\overline{X}-\mu)\sqrt{n}/S$作为解决这种情形的基础。

由本章第一节定理 1.1 知$T=(X-\mu)\sqrt{n}\big/s\sim t(n-1)$，则给定α(0<α<1)，查 t 分布分位表得$t_{\frac{\alpha}{2}}(n-1)$满足：$P\{|(\overline{X}-\mu)\sqrt{n}/S|<t_{\frac{\alpha}{2}}(n-1)\}=1-\alpha$，得

$$P\{\overline{X}-t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}<\mu<\overline{X}+t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}\}=1-\alpha \tag{4.3}$$

因此可取

$$(\overline{X}-t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}},\overline{X}+t_{\frac{\alpha}{2}}(n-1)\frac{S}{\sqrt{n}}) \tag{4.4}$$

作为μ的置信区间，对应的置信水平为(1−α)。

2 正态总体方差的区间估计

设正态总体 $X\sim N(\mu,\sigma^2)$，其中μ,σ^2均为未知，$X_1,\cdots,X_n$为其简单样本，给定α(0<α<1)，试对方差σ^2或标准差σ给出置信水平为(1−α)的置信区间。

由本章定理 1.1 知，令$\chi^2(n-1)=\dfrac{(n-1)S^2}{\sigma^2}$，则它是自由度为($n$−1)的卡方分布。若给定水平 (1−α)，可查$\chi^2(n-1)$表，存在两侧的分位数$\chi^2_{1-\frac{\alpha}{2}}(n-1)$及$\chi^2_{\frac{\alpha}{2}}(n-1)$满足：

$$P\{\chi^2(n-1)\geq\chi^2_{\frac{\alpha}{2}}(n-1)\}=\frac{\alpha}{2},\quad P\{\chi^2(n-1)<\chi^2_{1-\frac{\alpha}{2}}(n-1)\}=\frac{\alpha}{2}$$

则$P\{\chi^2_{1-\frac{\alpha}{2}}(n-1)<(n-1)S^2/\sigma^2<\chi^2_{\frac{\alpha}{2}}(n-1)\}=1-\alpha$，即：

$$P\{(n-1)S^2/\chi^2_{\frac{\alpha}{2}}(n-1)<\sigma^2<(n-1)S^2/\chi^2_{1-\frac{\alpha}{2}}(n-1)\}=1-\alpha \tag{4.5}$$

故可取

$$((n-1)S^2/\chi^2_{\frac{\alpha}{2}}(n-1),(n-1)S^2/\chi^2_{1-\frac{\alpha}{2}}(n-1)) \tag{4.6}$$

作为σ^2的置信区间。取

$$(\sqrt{n-1}S/\sqrt{\chi^2_{\frac{\alpha}{2}}(n-1)},\sqrt{n-1}S/\sqrt{\chi^2_{1-\frac{\alpha}{2}}(n-1)}) \tag{4.6a}$$

作为σ的置信区间。

3 两个正态总体方差比的区间估计

设两个正态总体 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，其中$\mu_1,\sigma_1^2,\mu_2,\sigma_2^2$为未知参数，$X_1,\cdots,X_{n_1}$ 与 $Y_1,\cdots,Y_{n_2}$ 分别是 X 与 Y 的样本，记 $S_1^2=\sum_{i=1}^{n_1}(X_i-\overline{X})^2/(n_1-1)$ 与 $S_2^2=\sum_{i=1}^{n_2}(Y_i-\overline{Y})^2/(n_2-1)$ 分别为它们的样本方差。给定置信水平(1–α)，试求方差比σ_1^2/σ_2^2的置信区间。

解：注意到$(n_1-1)S_1^2/\sigma_1^2$与$(n_2-1)S_2^2/\sigma_2^2$分别服从自由度为(n_1–1)和(n_2–1)的卡方分布，且相互独立（由取样相互独立），于是由F分布定义知

$$F=\frac{(n_2-1)S_2^2/\sigma_2^2(n_2-1)}{(n_1-1)S_1^2/\sigma_1^2(n_1-1)}=\frac{S_2^2/\sigma_2^2}{S_1^2/\sigma_1^2}=\frac{\sigma_1^2/\sigma_2^2}{S_1^2/S_2^2} \tag{4.7}$$

服从自由度为(n_2–1, n_1–1)的F分布。

对给定α，查F分布表，可得：

$$P\{F>F_{\frac{\alpha}{2}}(n_2-1,n_1-1)\}=P\{F<F_{1-\frac{\alpha}{2}}(n_2-1,n_1-1)\}=\frac{\alpha}{2}$$

则$P\{F_{1-\frac{\alpha}{2}}(n_2-1,n_1-1)<F<F_{\frac{\alpha}{2}}(n_2-1,n_1-1)\}=1-\alpha$，即得

$$P\{F_{1-\frac{\alpha}{2}}(n_2-1,n_1-1)\frac{S_1^2}{S_2^2}<\frac{\sigma_1^2}{\sigma_2^2}<F_{\frac{\alpha}{2}}(n_2-1,n_1-1)\frac{S_1^2}{S_2^2}\}=1-\alpha \tag{4.8}$$

故得σ_1^2/σ_2^2的置信水平为(1–α)的置信区间为：

$$(F_{1-\frac{\alpha}{2}}(n_2-1,n_1-1)\frac{S_1^2}{S_2^2},F_{\frac{\alpha}{2}}(n_2-1,n_1-1)\frac{S_1^2}{S_2^2}) \tag{4.9}$$

请读者思考：设总体 $X\sim N(\mu_1,\sigma_1^2)$，$Y\sim N(\mu_2,\sigma_2^2)$，若 σ_1^2,σ_2^2 已知，而 μ_1,μ_2 未知，给定α(0<α<1)，试给出均值差 $\mu_1-\mu_2$ 的区间估计。

4 大样本对总体均值的区间估计

以下讨论总体 $X\sim F(x)$是一般分布，但 $EX^4<\infty$ 存在，且均值 $EX=\theta$，方差 $DX=\sigma^2$ 均未知时，用大样本 $X_1,\cdots,X_n$（即 n 充分大）对总体均值的区间估计问题。

由本章第二节矩估计方法知，可用样本均值 $\overline{X}$ 作为总体均值 θ 的点估计，且 $\overline{X}$ 是 θ 的无偏估计与相合估计，即 $E\overline{X}=\theta$，且 $\lim\limits_{n\to\infty}\overline{X}\overset{p}{=}\theta$。又 $D\overline{X}=\sigma^2/n$，故由中心极限定理知，当 $n\to\infty$ 时，$(\overline{X}-\theta)\sqrt{n}/\sigma$ 的分布函数趋向标准正态概率分布 $N(0,1^2)$。另一方面，注意到 $ES^2=\sigma^2$，在 $EX^4<\infty$ 存在的条件下，可以证明当 $n\to\infty$ 时，$DS^2\to 0$，见[3]第二册。故由大数定律有：$\lim\limits_{n\to\infty}S^2\overset{p}{=}\sigma^2$。这样，在 $n\to\infty$ 时统计量 $(\overline{X}-\theta)\sqrt{n}/\sigma$ 与统计量 $(\overline{X}-\theta)\sqrt{n}/S$ 的极限分布均是标准正态分布 $N(0,1^2)$。即当 n 充分大时

$$U=(\overline{X}-\theta)\sqrt{n}/S$$

渐近标准正态分布。给定置信水平α，查标准正态分位数 $u_{\frac{\alpha}{2}}$，在 n 充分大时，有

$$P\{|\overline{X}-\theta|\sqrt{n}/S<u_{\frac{\alpha}{2}}\}\approx 1-\alpha$$

即

$$P\{\overline{X}-u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}<\theta<\overline{X}+u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}\}\approx 1-\alpha$$

于是 θ 的水平（1–α）的置信区间为：

$$(\overline{X}-u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}},\overline{X}+u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}) \tag{4.10}$$

例 4.2 从一大批钢筋中随机抽取 100 进行测试，结果有 4 件次品。试以水平为 95%的概率估计整批钢筋次品率的范围（置信区间）。

解：设总体 $X\sim B(1,p)$为二点分布，次品率为 p。X=0 表合格品，X=1 表次品。由题意样本容量 n=100，次品数 k=4，取 $S=\left(\left(\frac{k}{n}\right)\left(\frac{n-k}{n}\right)\right)^{\frac{1}{2}}=\sqrt{0.04\times 0.96}$，而后利用(4.10)式可

得：

$$\frac{k}{n}-u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}=0.04-1.96\times\sqrt{0.04\times0.96}/10=0.002$$
$$\frac{k}{n}+u_{\frac{\alpha}{2}}\frac{S}{\sqrt{n}}=0.04+1.96\times\sqrt{0.04\times0.96}/10=0.078$$

故整批钢筋次品率的置信区间为(0.002, 0.078)。

5. 贝叶斯区间估计

设总体 $X\sim f(x,\theta)$，其中 $\theta\in\Theta$ 为未知参数，先验分布为 $h(\theta)$，给定样本 X_1, ⋯, X_n，θ 的后验分布为 $h(\theta|X_1,\cdots,X_n)$。当 $(X_1,\cdots,X_n)=(x_1,\cdots,x_n)$ 时，对给定水平 1–α，若区间 $[\tilde{\theta}_1(x),\tilde{\theta}_2(x)]$ 满足：

$$\int_{\tilde{\theta}_1(x)}^{\tilde{\theta}_2(x)}h(\theta\,|\,x)d\theta=1-\alpha$$

则称 $[\tilde{\theta}_1(x),\tilde{\theta}_2(x)]$ 为置信水平 1–α参数 θ 的贝叶斯区间估计。

例 4.3 设总体 $X\sim N(\mu,\sigma^2)$，μ 的先验分布为 $N(\mu_0,\sigma_0^2)$，其中 σ, μ_0,σ_0^2 均为已知常数，$X_1,\cdots,X_n$ 为 X 的样本，试求 μ 的 *Bayes* 区间估计。

解：有前面例，可知 μ 的后验分布 $h(\theta\,|\,x_1,\cdots,x_n)\sim N(a,b^2)$，其中

$$a=(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})^{-1}(n\bar{x}/\sigma^2+\mu_0/\sigma_0^2)，\quad b^2=(\frac{n}{\sigma^2}+\frac{1}{\sigma_0^2})^{-1}$$

由 $h(\theta\,|\,x_1,\cdots,x_n)\sim N(a,b^2)$，则 $\frac{\mu-a}{b}$ 关于 x 的分布为 $N(0,1^2)$，故而

$$P\{-u_{\alpha/2}<\frac{\mu-a}{b}<u_{\alpha/2}\,|\,x\}=1-\alpha$$

于是 μ 的贝叶斯（1–α）水平的区间估计为：$[a-u_{\alpha/2}b,a+u_{\alpha/2}b]$。

置信区间是下章假设检验的基础，同时有广泛的应用背景。

§5 非参数估计

以上讨论参数估计问题，本小节讨论非参数估计问题，主要讨论总体分布函数的估计问题。

设总体 $X\sim F(x)$, 其样本为 $X_1,X_2\cdots,X_n$， $\forall x\in R$，记 $\widetilde{F}_n(x)\hat{=}\frac{1}{n}\sum_{k=1}^{n}I(X_k\le x)$，称 $\widetilde{F}_n(x)$ 为 X 的经验分布，若用 $\widetilde{F}_n(x)$ 作为总体分布函数的估计，则由第三、四章内容可知：$\forall x\in R$ 有

（1） $E\widetilde{F}_n(x)=F(x)$， $D\widetilde{F}_n(x)=F(x)(1-F(x))/n$.

(2) $\forall\varepsilon>0$ ，$\lim\limits_{n\to\infty}P(|\widetilde{F}_n(x)-F(x)|<\varepsilon)=1$。

事实上，有更强的结果：

(3)* $P(\lim\limits_{n\to\infty}\widetilde{F}_n(x)=F(x))=1$.

可见 $\widetilde{F}_n(x)$ 是 $F(x)$的无偏估计与（强）相合估计。（所指强相合，即当 $n\to\infty$ 时，$\widetilde{F}_n(x)$ 以概率 1 收敛到 $F(x)$ ）。

练习题

10.1 设 $x_1,\cdots x_n$ 是来自正态分布 $N(\mu,\sigma^2)$ 的样本值，μ 已知，求 σ^2 的极大似然估计

量。

10.2 设$x_1,\cdots x_n$是来自正态分布$N(\mu,1)$ 的样本值，求μ的极大似然估计量。

10.3 设X服从区间$[0,\lambda]$（$\lambda>0$）上的均匀分布，λ是未知参数，而$x_1,\cdots x_n$是X的样本值，试求出λ的极大似然估计量和矩估计量。

10.4 设$X_1,\cdots X_n$是来自均匀分布$U(\theta,2\theta)$的样本, 试求出θ的最大似然估计量和矩估计量。

10.5 设X服从区间$[a,b]$ 上的均匀分布，这里a,b是两个未知参数。若$x_1,\cdots x_n$(不全相等)是X的样本值，试求出a,b的最大似然估计量。

10.6 设X~$Po(\lambda)$，参数λ有先验密度$h(\lambda)$=$e^{-\lambda}I_{(\lambda\geq 0)}$。试分别求出$\lambda$的贝叶斯最大后验估计及贝叶斯条件期望估计。

10.7 设$X_1,\cdots X_n$是来自指数分布的样本。分布中的参数λ有先验密度$h(\lambda)=e^{-\lambda}I_{(\lambda>0)}$。试分别求出$\lambda$的贝叶斯最大后验估计及贝叶斯条件期望估计。

10.8 设X是来自二项分布$B(n,p)$的样本, n已知，p为未知参数。证明：对任何常数c,d，可找到p的先验分布。使p的贝叶斯估计为$(X+c)/(n+d)$。

10.9 设X是来自泊松分布$Po(\lambda)$的样本。(*a*) 证明：$g(\lambda)=e^{-2\lambda}$的唯一的无偏估计$\hat{\theta}(X)$为：$\hat{\theta}(X)=1$当X为偶数，$\hat{\theta}(X)=0$当X为奇数。(*b*) 你认为 (*a*) 中的估计是否合理？如不合理，试提出一个合理的估计。

10.10 设$X_1,\cdots X_n$是来自正态总体$N(a,\sigma^2)$的样本，则已知$\hat{\theta}_1(x)=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2$为$\sigma^2$之一无偏估计。证明：$\hat{\theta}_2=\frac{n-1}{n+1}\hat{\theta}_1$虽非$\sigma^2$的无偏估计，但$\hat{\theta}_2$的均方误差较小，即：$E(\hat{\theta}_2-\sigma^2)^2<E(\hat{\theta}_1-\sigma^2)^2$。本题说明：无偏估计不一定是最好的

选择。

10.11 设 $X_1,\cdots X_n$ 是来自某一个具有均值 θ 而方差有限的总体中抽出的样本。证明：对任何常数 $c_1,\cdots,c_n$，只要 $\sum_{i=1}^{n} c_i=1$，则 $\sum_{i=1}^{n} c_i X_i$ 必是 θ 的无偏估计。但是，只有在 $c_1=\cdots=c_n=1/n$ 时方差达到最小（指在上述形式的估计类中达到最小。实际可以证明：$\overline{X}$ 在 θ 的一切无偏估计类中也达到最小。）

10.12 设 $X_1,\cdots,X_n$ 是来自均匀分布 $U(0,\theta)$ 的样本。证明：对任给的 $1-\alpha$ $(0<1-\alpha<1)$，可找到常数 c_n，使 $[\max(X_1,\cdots,X_n), c_n\max(X_1,\cdots,X_n)]$ 为 θ 的一个置信系数 $1-\alpha$ 的区间估计。

10.13 设 $X_1,\cdots,X_n$ 和 $Y_1,\cdots,Y_m$ 分别是来自正态总体 $N(\theta,\sigma_1{}^2)$ 和 $N(\theta,\sigma_2{}^2)$ 的样本，$\sigma_1{}^2$ 和 $\sigma_2{}^2$ 都已知。(*a*)找常数 c,d，使 $\hat{\theta}=c\overline{X}+d\overline{Y}$ 为 θ 的无偏估计。并使其方差最小（在所有形如 $c\overline{X}+d\overline{Y}$ 的无偏估计类中最小）。(*b*)基于 $\hat{\theta}$，作出 θ 的置信系数为 $1-\alpha$ 的置信区间。

10.14 测量铝的比重 16 次，测得 $\overline{x}=2.705, S=0.029$, 试求铝的比重的置信区间（设测量值服从正态分布，置信度为 0.95）。

10.15 设 $X\sim N(\mu,\sigma^2)$，$x_1,\cdots x_n$ 是其样本值。如果 σ^2 已知，问：n 取多大时方能保证 μ 的置信度为 0.95 的置信区间的长度不大于给定的 L?

10.16 设 $X_1,\cdots,X_n$ 是来自正态总体 $N(\theta,\sigma^2)$ 的样本，θ 和 σ^2 都未知。证明：$\overline{X}$ 仍为 θ 的 *MVU* 估计。

10.17 设总体 X 为 *d.r.v.*，其分布律为 $P(X=-1)=(1-\theta)/2, P(X=0)=1/2,$ $P(X=1)=\theta/2$，$X_1,\cdots,X_n$ 为其样本。

（1）求θ的 *MLE*（极大似然估计）$\tilde{\theta}_1$，问 $\tilde{\theta}_1$ 是否是θ的无偏估计；

（2）求θ的矩估计$\widetilde{\theta}_2$；

（3）计算θ的无偏估计的方差 C—R 下界。

10.18 设总体 X 的 $p.d.f.$为 $f(x,\theta)=\frac{1}{\theta}x^{\frac{1}{\theta}-1}I_{(0<x<1)}$ $\theta>0$，$X_1,\cdots,X_n$ 为其样本。

（1）求θ的 MLE $\widetilde{\theta}_1$，$E\widetilde{\theta}_1$及 $D\widetilde{\theta}_1$；

（2）求θ的最小方差无偏估计$\widetilde{\theta}_2$。

10.19 设总体 X 是服从参数为 p 的负二项分布，即

$$P(X=k)=C_{k-1}^{r-1}(1-p)^{k-r}p^r \quad (k\geq r, r\geq 1\text{为整数})$$

$X_1,\cdots,X_n$ 为其样本。

（1）求 p 的 MLE $\widetilde{p}_1$；

（2）求 p 的矩估计$\widetilde{p}_2$。

10.20 设总体 $X\sim U(\theta_1,\theta_2)\,(\theta_1<\theta_2)$，$X_1,\cdots,X_n$ 为其样本，$X_{(1)},\cdots,X_{(n)}$ 为其顺序统计量。

（1）求θ_1与θ_2的最小方差无偏估计；

（2）取$\theta_2=2\theta_1$，求θ_1的 MLE，问它是否是θ_1的无偏估计，说明理由；

（3）当$\theta_2=2\theta_1$时，令$T=(n+1)(2X_{(n)}+X_{(1)})/(5n+4)$，试证 T 是θ_1的无偏估计，且$DT\leq D\widetilde{\theta}_1$。

10.21 设总体 $X\sim U(k\theta, (k+1)\theta)$，($k$>0 已知)。$X_1,\cdots,X_n$ 为其样本，$X_{(1)},\cdots,X_{(n)}$ 为其顺序统计量。

（1）求θ的 MLE，记为$\widetilde{\theta}$；

（2）令$\hat{\theta}=X_{(1)}/k$，$T_\alpha=\alpha\widetilde{\theta}+(1-\alpha)\hat{\theta}$，求使$DT_\alpha$达最小的α*，并计算$DT_{\alpha^*}/D\widetilde{\theta}$。

10.22 设总体 X 的 $p.d.f.$为 $f(x,\mu,\sigma)=(1/2\sigma)\exp\{-|x-\mu|/\sigma\}$，$\mu\in R,\sigma>0$，$X_1,\cdots,X_n$ 为其样本。

（1）求 μ,σ 的矩估计；

（2）求 μ,σ 的极大似然估计。

10.23 设 $X_1,\cdots,X_n$ *i.i.d.*，EX_1=μ，DX_1<∞。证明 $\tilde{\mu}=\frac{2}{n(n+1)}\sum_{i=1}^{n}iX_i$ 是μ的相合估计。

10.24 设 $X_1,\cdots,X_n$ 是来自总体 $X\sim U(0, n\lambda)$的一个样本， $X_{(1)},\cdots,X_{(n)}$ 为其顺序统计量。

（1）证明 $n\lambda-X_{(1)}$ 的分布收敛于 $\exp(1/\lambda)$；

（2）用（1）中的结果给出λ的一个相合估计。

10.25 设总体 X 的 *p.d.f.*为 $f(x,\theta)=\theta(1+\theta)x^{\theta-1}(1-x)I_{(0<x<1)},\theta>0$，$X_1,\cdots,X_n$ 为其样本。

（1）求θ的矩估计 $\tilde{\theta}_n$；

（2）求 $\sqrt{n}(\tilde{\theta}_n-E\tilde{\theta}_n)$ 的渐近分布；

（3）$\tilde{\theta}_n$ 是否是θ的渐近有效估计？

10.26 从一批电子元件中抽取 100 件，若抽取的元件的平均强度为 1000，标准差为 40，假设该批元件强度 $X\sim N(\mu,\sigma^2)$，试求μ的置信区间（设α=0.05）。

10.27 设某厂每天生产的一批钢筋强度 $X\sim N(\mu,\sigma^2)$。现从中抽取 20 件，测得抗拉强度为：45.20，44.90，45.11，45.20，45.54，45.38，44.77，45.35，45.15，45.11，45.00，45.61，44.88，45.27，45.38，45.46，45.27，45.23，44.96，45.35。给定α=0.05，试求μ与 σ 的置信区间。

10.28 设某一钢厂生产的大批量钢材，每根强度 $X\sim N(\mu,\sigma^2)$，其中 $\sigma=10(kg)$ 为已知。均值 μ 为未知参数。但由以往生产的记录知，每天生产的每批钢材的期望强度 $\mu\sim N(1150,20^2)$。现从某一天生产的一批钢材中任取 5 根，测得每根钢材的强度为 1145，1150，1147，1151，1155。试用贝叶斯条件期望估计法，对该批钢材的条件期望强度作出估计。

10.29 设总体 $X\sim B(1,p)$，其中 p 为未知参数， $p\in[0,1]$。$X_1,\cdots,X_n$ 为 X 的样本，记 $X=\sum_{i=1}^{n}X_i$，若 p 的先验分布 $h(p)$ 取为参数是 (α,β) 的贝塔分布，即 $h(p)\sim B_e(\alpha,\beta)$（详见第二章定义）。若试验结果为 $X=r\ (0\le r\le n)$，试求 p 的贝叶斯最大后验估计值。

10.30 设总体$X \sim Po(\theta)$（即参数为θ的泊松分布），$\theta > 0$为未知参数。按已知历史资料，选取θ的先验分布$h(\theta)$为参数是(α, r)的伽玛分布（详见第二章伽玛分布的定义）。现有样本$X_1, \cdots, X_n$取值为$x_1, \cdots, x_n$，记$\bar{x} = \sum_{i=1}^{n} x_i / n$，试求$\theta$的最大后验估计值。

10.31 定数截尾寿命试验。设总体$X \sim F(x, \theta)$，密度函数为$f(x, \theta), \theta \in \Theta$，$X \geq 0$为非负$r.v.$。若$X$表示一批元件的工作寿命，现从中抽取$n$个样本进行寿命测试。为克服获得完全的寿命数据所需时间过长的缺陷，采取定个数截尾寿命试验如下：即事先指定一个数k（$1 \leq k \leq n$），设抽取的n个元件从t=0开始工作，当实验进行到有k个样品失效就停止观测。这时，仍有$(n-k)$个样品尚未失效。若记前k个先后失效时间为$X_{(1)}, X_{(2)}, \cdots, X_{(k)}$。此$X_{(1)}, X_{(2)}, \cdots, X_{(k)}$为$n$个样品定数截尾寿命试验数据。记$T_{n,k} = \sum_{i=1}^{k} X_{(i)} + (n-k) X_{(k)}$，$\tilde{\theta}_{n,k} = T_{n,k} / k \quad (1 \leq k \leq n)$。若$f(x, \theta) = \theta^{-1} e^{-\theta^{-1} x} I_{(x>0)}$，$\theta \in \Theta$未知。

（1）给定$X_{(1)}, X_{(2)}, \cdots, X_{(k)}$，试求$\theta$的极大似然估计量$\tilde{\theta}_L$；

（2）证明$\tilde{\theta}_{n,k}$是θ的无偏估计$(1 \leq k \leq n)$；

（3）证明$\tilde{\theta}_{n,k}$是θ的相合估计$(1 \leq k \leq n)$。

10.32 同上题定数截尾寿命试验，但总体X是参数为(α, γ)的伽玛分布，即$X \sim G(\alpha, \gamma)$。若$\gamma = 2$已知，取n个样本，给定$X_{(1)}, X_{(2)}, \cdots, X_{(k)}$，试求$\alpha$的极大似然估计量$\tilde{\alpha}_L$。

第十一章 假设检验

本章简要介绍假设检验的基本概念与推断的直观依据，介绍简单的假设检验的基本方法和一般步骤，重要参数的检验方法及分布函数的检验方法。

§1 问题的提法与基本概念

所谓假设检验是指根据样本提供的信息来推断（检验）总体特性的某些假设是否可信（是否成立）。

例 1.1 某厂生产一大批元件，按质量标准规定，次品率要低于 1%，并须经抽样检验合格才能出厂，现从其中任抽取 5 件，发现 5 件中含有次品，问这批产品能否出厂？

从直观上判断，这批元件是不能出厂的。那么理由何在？记这批元件的次品率为 p，若记假设（或命题）$H_0=\{p<0.01\}$，X 为抽取的次品个数。则问题可转化为：如何根据抽样结果检验假设 H_0 是否成立呢？可以这样来推断：如果假设 H_0 成立，即 $p<0.01$，那么抽取 5 个抽到次品的概率为 $P(X\geq 1)=1-P(X=0)<1-(0.99)^5<0.05$，即事件$(X\geq 1)$是一小概率事件。而小概率事件在一次试验中是难以发生的。但一次试验结果竟然发生，于是我们可以有把握断言，原来假设 H_0 难以置信，应拒绝 H_0 这一原先假设。这就是根据样本结果检验总体原先假设 H_0 是否成立的直观依据。称这种推断依据为小概率事件推断原理。

显然，抽取的样品中次品的个数越大，假设 H_0 就越不可信，就越应该被拒绝。通常对于一次的简单抽样，对样品中的次品数事先要规定一个临界值 K。例如上例中，若规定 $K=0$，当$(X=0)$时接受假设 H_0，当$(X>0)$时拒绝 H_0 假设。

一般地，若厂方规定每次抽取 n 个元件进行检验，仍记 X 为抽验中的次品数，规定临界值为 K。当$(X\leq K)$时，接受原假设 H_0，即认为 H_0 成立，允许该批元件出厂；当$(X>K)$时，拒绝 H_0 假设，即认为 H_0 不成立，不允许该批元件出厂。称$(X\leq K)$为接受区域（即接受 H_0 假设的区域），$(X>K)$为拒绝域。

给定 H_0 与 n，如何确定拒绝域或临界值 K 呢？确定 K 的基本依据是：当 H_0 成立时，选取 K，$(X>K)$应是一个小概率事件。确切地说，预先给定小概率α（$0<\alpha<1$，一般取为 0.05 或 0.01 等），选取 K 满足：当 H_0 成立时，$P(X>K)\leq\alpha$，即确定临界值 K 应保证 H_0 成立时，X 落在拒绝域的概率是不超过α的小概率。

由上例可知，所谓假设检验就是，根据研究问题的需要，对总体的某些性质提出假设，如假设 H_0：$p<0.01$，称 H_0 假设为原假设或零假设。为判断原假设 H_0 是否成立，采用类似于“反证法”的思想方法。我们先假定这个“假设 H_0”成立，然而 H_0 是否成立要由试验结果来验证：如果由假定 H_0 成立条件下的某一个小概率事件，竟然在一次试验

中发生，我们就有理由拒绝假设 H_0（起码不能轻易接受 H_0），如果必须在拒绝与接受之间二者选择之一，那么此时选择拒绝 H_0 比选择接受 H_0 更合理。相反，若在一次试验中发生的是一大概率事件（在原假设 H_0 成立的条件下），那么，这时选择接受原假设 H_0 比拒绝 H_0 更合理更明智。这种推断方法我们称之为小概率事件统计推断原理。

那么，事件的概率小到什么程度才算“小概率事件”呢？这要视实际具体问题的需要而定。一般把概率不超过 0.05 的事件当作“小概率事件”。例如，普通的产品抽验，通常取α=0.05，α=0.01；若对降落伞这一类产品通常要求α小于 0.001，而对飞机航天器就更要小的多了。

例 1.2 某奶粉厂用自动打包机打包，规定每包奶粉标准重为 500 克，已知每包重量服从正态分布，其均方差 $\sigma = 0.9$ 克长期保持不变。每天开工后需抽样检验打包机工作是否正常（即有无系统偏差）。某日开工后测得 9 包重量（单位：克）如下：

500.7，500.5，499.7，501.0，500.8，499.8，501.2，502.0，500.6

给定α=0.05。问该日打包机工作是否正常？

解：对抽样数据的初步直观分析：设每包重为 X，按规定，若打包机工作正常，则应有：$X \sim N(500, 0.9^2)$。那么在正常情况下，样本数据应该大体是：（i）数据应围绕 500 左右波动，且大于和小于 500 的数目应各占 1/2 左右；（ii）样本数应有约 95%左右落在区间 $(500 - 2\times 0.9, 500 + 2\times 0.9)$ 上。而抽样结果是：（i）9 个数据只有 2 个小于 500，而有 7 个大于 500；（ii）且 9 个数据已有一个落在区间 $(500 - 2\times 0.9, 500 + 2\times 0.9)$ 之外。初步印象，打包机的平均重量可能偏重！不正常，应进一步分析检验。以下介绍 U—检验法。

依题意 $X \sim N(\mu, \sigma^2)$，且 $\sigma = 0.9$ 已知。为检验打包机工作是否正常，我们先提出假设 H_0： $\mu = \mu_0 = 500$，然后用抽验结果检验 H_0 是否成立。若 H_0 成立的，则 $U = (\overline{X} - \mu_0)\sqrt{n} / \sigma \sim N(0,1^2)$，因此，可以用 U 的取值来判断 H_0 是否成立。给定α=0.05，取 $u_{\alpha/2} = u_{0.025}$=1.96 使 $P\{|U| < 1.96\} = 1 - \alpha = 0.95$。因此，若 H_0 成立，则 $\{|U| > 1.96\}$ 是小概率事件。而小概率事件在一次试验中基本不会发生。但是计算 $\overline{X}$ 的取值 $\bar{x} = 500.7$ 及统计量 U 的取值 $u = (\bar{x} - \mu_0)\sqrt{n} / \sigma = (500.7 - 500)\times\sqrt{9} / 0.9 = 2.33 > 1.96$，这意味着小概率事件 $\{|U| > 1.96\}$ 竟然在一次试验中发生了，这是不合理的。究其源是因为原假设 H_0 不可信。因而应拒绝 H_0，即认为不能讲该日打包机工作正常，应先调整之后再生产。

设总体 $X \sim N(\mu,\sigma^2)$，σ^2 已知，检验均值 μ 的一般步骤：

（1）根据问题要求，提出原假设 H_0：$\mu = \mu_0$，对立假设 H_1：$\mu \neq \mu_0$；

（2）根据 H_0，选取统计量 $U = (\overline{X} - \mu_0)\sqrt{n}/\sigma$，在 H_0 成立条件下 $U \sim N(0,1^2)$；

（3）给定α（称α为显著水平），查表定出临界值 $u_{\alpha/2}$，从而使 H_0 成立时，事件 $\{|U| > u_{\alpha/2}\}$ 是一小概率事件；

（4）计算 $\overline{X}$ 的取值 $\bar{x}$ 及统计量 U 的取值 $u = (\bar{x} - \mu_0)\sqrt{n}/\sigma$，当 $|u| > u_{\alpha/2}$ 时，拒绝 H_0，当 $|u| < u_{\alpha/2}$ 时，接受 H_0。

通过以上诸例，可初步了解到假设检验问题的提法，以及用于检验假设的直观依据即小概率事件统计推断原理。在此，对若干常用名词与概念补充如下：

（1）原假设与对立假设

在假设检验中，常把一个被检验的“重要” 假设叫做原假设，用 H_0 表示，或叫做零假设。而与原假设不同的另一假设，称为备择假设，用 H_1 表示，有时也叫做对立假设。如例 1.2 中，原假设为 H_0：$\mu = \mu_0$=100，而对立假设或备择假设为 H_1：$\mu \neq \mu_0$。

（2）检验统计量，接受域，拒绝域

为检验一个假设 H_0 所选择的样本统计量称为检验统计量。使 H_0 得到接受的所有样本点 $X_1,\cdots X_n$ 所在区域全体，称为该检验的接受域，记为 W_0。而使 H_0 被拒绝的所有样本点 $X_1,\cdots X_n$ 所在区域全体，称为该检验的拒绝域，记为 W_1。如例 1.2 中，H_0 的接受域为：W_0= $(\mu_0 - u_{\alpha/2}\sigma/\sqrt{n}, \mu_0 + u_{\alpha/2}\sigma/\sqrt{n})$，$H_0$ 的拒绝域为：W_1=$(-\infty, \mu_0 - u_{\alpha/2}\sigma/\sqrt{n}) \bigcup (\mu_0 + u_{\alpha/2}\sigma/\sqrt{n}, +\infty)$，点 $\mu_0 \pm u_{\alpha/2}\sigma/\sqrt{n}$ 称为临界点。

（3）单式检验，复式检验与序贯抽样检验

根据抽取一回容量为 n 的样本就作出或者接受 H_0，或者拒绝 H_0 的决定的检验称为

单式检验。对复式检验，以例 1.1 的情形为例，若要检验 H_0：$p<0.01$ 是否成立，规定抽样方案如下：第一回抽取 5 个，记其中红球个数为 X_1，若 $X_1=0$，则接受 H_0；若 $X_1\geq 2$，则拒绝 H_0；若 $X_1=1$，则再抽第二回，例如再抽取 10 个。记 X_2 为其中红球数，若 $X_2=0$，则接受 H_0，否则拒绝 H_0。这种抽验方式称为复式抽验。而第二章练习题 2.31 题所叙述的抽样检验就是一种序贯抽验。

§2 两类错误与功效函数

1. 检验法

由前面的例子可知，在假设检验中，对给定的假设 H_0 作出接受或拒绝的决断，这要根据样本的取值情况事先规定一个规则，我们称之为检验法。确切地说，记 S 为样本取值 $x=(x_1,\cdots,x_n)$ 全体构成的集合（样本空间），那么一个检验Φ是指对 S 的一个划分：W_0，W_1，满足：$W_0\bigcup W_1=S, W_0\bigcap W_1=\phi$。当样本值 $x=(x_1,\cdots,x_n)\in W_0$ 时就接受 H_0（拒绝 H_1）；当样本值 $x=(x_1,\cdots,x_n)\in W_1$ 时就拒绝 H_0（接受 H_1）。简记检验Φ=(W_0, W_1)。

例如例 1.2 中，H_0：$\mu=\mu_0$，一个检验Φ：$W_0=(\mu_0-u_{\alpha/2}\sigma/\sqrt{n},\mu_0+u_{\alpha/2}\sigma/\sqrt{n})$，$W_1=(-\infty,\mu_0-u_{\alpha/2}\sigma/\sqrt{n})\bigcup(\mu_0+u_{\alpha/2}\sigma/\sqrt{n},+\infty)$。

2. 两类错误

由于假设检验是根据一次抽样所得的检验统计量的取值是落在拒绝域或接受域而作出决定是否拒绝与接受 H_0。这种由部分信息判断总体的结果有可能发生两种错误：（1）当 H_0 成立时，而经检验被拒绝的错误，称为第一类错误，犯第一类错误的概率记为 α；（2）当 H_0 不成立时，而经检验被接受的错误称为第二类错误，犯第二类错误的概率记为 β。

如何求其 α 与 β 呢？以例 1.2 说明。当 H_0：$\mu=\mu_0=500$ 成立时，若取 $\alpha=0.05$，则在 H_0 成立条件下统计量落在拒绝域 $\{|U|>u_{0.025}\}$ 的概率即为 α；当 $\mu\neq\mu_0=500$ 时，即 H_0 不成立时，不妨设 $\mu=\mu_1\neq\mu_0$ 时，求统计量 U 落在接受域的概率。此时，$\overline{X}\sim N(\mu_1,\sigma^2/n)$，于是，$\overline{X}$ 落在接受域的概率为：

$$\beta = P_{\mu_1}(\mu_0 - u_{\alpha/2}\sigma/\sqrt{n} < \overline{X} < \mu_0 + u_{\alpha/2}\sigma/\sqrt{n})$$
$$= (\sqrt{2\pi}\sigma/\sqrt{n})^{-1}\int_{\mu_0 - u_{\alpha/2}\sigma/\sqrt{n}}^{\mu_0 + u_{\alpha/2}\sigma/\sqrt{n}} \exp[-(u-\mu_1)^2 n/2\sigma^2]du$$

令 $t = (u-\mu_1)\sqrt{n}/\sigma$ 代入上式，得：

$$\beta = \Phi((\mu_0 - \mu_1)\sqrt{n}/\sigma + u_{\alpha/2}) - \Phi((\mu_0 - \mu_1)\sqrt{n}/\sigma - u_{\alpha/2}) \tag{2.1}$$

由 (2.1) 式易知，当 n 给定时，$|\mu_0 - \mu_1|$ 越大，则β越小。若给定 $\mu = \mu_1$ 固定，则当 $n \to \infty$ 时，$(\mu_0 - \mu_1)\sqrt{n}/\sigma$ 趋于正无穷(若 $\mu_0 > \mu_1$)或负无穷大（若 $\mu_0 < \mu_1$），故 $n \to \infty$ 时，β□0。

通常在检验一个假设 H_0（及 H_1）时，希望犯两类错误的概率都尽可能小，但这两者的要求往往是矛盾的。仍以例 1.2 为例来看。当犯第一类错误的概率 α 减小时，则由 $P(|U| > \mu_{\alpha/2}) = \alpha$ 知，对应 $u_{\alpha/2}$ 的就增大，又由(2.1)易看出，当 $u_{\alpha/2}$ 增大时，对应的 β（犯第二类错误的概率）就增大，反之亦然。因此需要建立适当的准则以权衡与协调二者的利害得失。下面仅介绍在假设检验中如何区分不同检验法的一个基本依据和一种较为流行的处理方法。

3. 功效函数（势函数）

同一原假设 H_0：$\theta \in \Theta_0$（其中 $\Theta_0 \subset \Theta$ 是 Θ 的非空真子集），可以用不同的检验法对 H_0 进行检验。那么，根据什么来区分不同检验法的优劣呢？引进以下定义。

定义 设对总体 X 的分布 $f(x,\theta)$ 提出原假设 H_0：$\theta \in \Theta_0$（其中 $\Theta_0 \subset \Theta$ 是 Θ 的非空真子集），对立假设 H_1：$\theta \in \Theta - \Theta_0$。记其样本为 $(X_1, \cdots, X_n)$，若采用检验Φ=(W_0, W_1) 对 H_0 进行检验，称

$$\beta_\Phi(\theta) = P_\theta((X_1, \cdots, X_n) \in W_1) \tag{2.2}$$

为检验Φ的功效函数（*power function*，又称为势函数）。

易知，当且仅当事件 $\{(X_1, \cdots, X_n) \in W_1\}$ 发生时，拒绝原假设 H_0，而这事件的概率显然与参数θ有关，即它是定义在参数空间 Θ 上的函数。

不难看出，当 $\theta \in \Theta_0$ 时，$\beta_\Phi(\theta)$ 表示犯第一类错误的概率；当 $\theta \in \Theta_1$ 时，$1 - \beta_\Phi(\theta)$

表示犯第二类错误的概率。换言之，当$\theta \in \Theta_0$时，要求$\beta_\Phi(\theta)$尽量小（即当H_0成立时，我们要求不要轻易拒绝它）；而当$\theta \in \Theta_1$时，要求$\beta_\Phi(\theta)$尽量大（即当H_0不成立时，希望能严格把好关，拒绝它）。这说明一个检验的功效函数确实反应该检验法的优劣水平。

如何协调和解决两类错误的利害得失呢？目前在理论上较为流行的办法是沿用奈曼—皮尔逊（*Neyman* 和 *Pearson*）提出的假设检验的基本思想，就是首先确保犯第一类错误的概率限制在不超过一给定值α的前提下，然后寻找使犯第二类错误的概率尽可能小的检验。在这种准则下，寻找一个好的检验法，就是事先约定的一个小的正数α（0<α<1），在满足：

$$\beta_\Phi(\theta) = P_\theta((X_1,\cdots,X_n) \in W_1) \le \alpha \qquad \forall \theta \in \Theta_0 \tag{2.3}$$

的检验法中，寻找检验Φ，使得当$\theta \in \Theta_1$时，$\beta_\Phi(\theta)$尽可能的大。

在$\theta \in \Theta_0$时，满足$\beta_\Phi(\theta) \le \alpha$的检验Φ称为水平为α的检验。简记为$(\alpha, \Theta_0, \Theta_1)$检验。称$\alpha$为检验的显著性水平（*evidence level*），简称为水平。

§3 常用的参数检验

本节将讨论几个常用的参数检验，以及在奈曼－皮尔逊假设检验的基本准则下，如何选择与构造在直观上看来较为合理的检验。

1. 正态总体均值的检验及推广

设总体$X \sim N(\mu, \sigma^2)$，对均值μ的假设检验仅介绍以下三种典型形式。

（1°）H_0： $\mu = \mu_0$，H_1： $\mu \ne \mu_0$；（双侧（边）假设检验）

（2°）H_0'： $\mu \ge \mu_0$，H_1'： $\mu < \mu_0$；（单侧假设检验）

（3°）H_0''： $\mu \le \mu_0$，H_1''： $\mu > \mu_0$。（单侧假设检验）

分别讨论：

（1）方差σ^2已知时，检验H_0、H_0'、H_0'' 。

对检验H_0： $\mu=\mu_0$， H_1： $\mu\neq\mu_0$，可见例 1.2，已作讨论，这里不重复。在此例 1.2 中所用的检验统计量为$U=(\bar{X}-\mu_0)\sqrt{n}/\sigma$，因此称之为 U—检验法。

这里只对检验H_0'： $\mu\geq\mu_0$， H_1'： $\mu<\mu_0$作讨论。记$\bar{X}$为样本均值，以$\bar{X}$作为μ的估计。故若$\bar{X}$越大，直观上与原假设H_0'：$\mu\geq\mu_0$越吻合；反之，若$\bar{X}$越小，则与H_1'：$\mu<\mu_0$更贴近。因此，在直觉上一个较为合理的检验Φ应是：选取检验量为：$U=(\bar{X}-\mu_0)\sqrt{n}/\sigma$，对给定的α，确定$u_\alpha$，务请注意在这种情形，应使$P(U<-u_\alpha)=\alpha$ 。于是当H_0'成立时，事件$(U<-u_\alpha)=((\bar{X}-\mu_0)\sqrt{n}/\sigma<-u_\alpha)$应是小概率事件。因此，当$\{(\bar{X}-\mu_0)\sqrt{n}/\sigma<-u_\alpha\}$发生时，应拒绝$H_0'$；反之，则接受$H_0'$。所以Φ的接受域与拒绝域应分别为：

$$\Phi：W_0=(\mu_0-u_\alpha\sigma/\sqrt{n},+\infty),\quad W_1=(-\infty,\mu_0-u_\alpha\sigma/\sqrt{n}),$$

称上述检验法为单边—检验。而例 1.2 中的检验法称为双边 U—检验法。

在许多实际检验问题中，会遇到原假设$H_0{}':\mu\geq\mu_0$的单侧（边）假设检验问题。例如，某工厂每天生产的一批钢筋的质量指标之一是它的平均抗拉强度μ越大越好。按出厂的质量标准，要求$\mu\geq\mu_0$才可出厂，因此提出原假设为 $H_0{}':\mu\geq\mu_0$。这时制定的抽样检验法，自然希望对$H_0{}':\mu\geq\mu_0$成立的钢筋，经检验尽可能通过（接受$H_0{}'$）。而且μ越大，通过检验后接受的概率应越大越好；反之，对于$\mu<\mu_0$的钢筋，经检验不让轻易通过（即拒绝$H_0{}'$）的概率应越大越好。这时用原假设$H_0{}':\mu\geq\mu_0$，采用上述单侧拒绝

域 $W_1{}' = (-\infty, \mu_0 - \mu_\alpha \sigma \big/ \sqrt{n})$ 就能达到这个目的。这要比用双侧拒绝域 $W_1 = (-\infty, \mu_0 - \mu_{\alpha/2} \sigma \big/ \sqrt{n}) \cup (\mu_0 + \mu_{\alpha/2} \sigma \big/ \sqrt{n}, \infty)$ 的检验法更为优良，即更为合理。

请读者举出单侧假设检验的其它例子。

对于检验 H_0''： $\mu \le \mu_0$ ， H_1''： $\mu > \mu_0$ 的情形的检验与上述类似，从略。读者可作为练习，讨论并给出检验 H_0'' 的较为合理的检验法。对于这种单侧假设检验 $H_0{}'' : \mu \le \mu_0$ ，在质量检验中也会遇到。例如，衡量大批量生产的电子元件的质量指标是每批元件的次品率 p $(0 \le p \le 1)$，显而易见，p 越小，质量越好。这时可采用单侧假设检验 $H_0{}'' : p \le p_0$ 较为合适。

对于σ已知，对 μ 的其它形式的假设检验问题可参照以上三种典型，给出适当的检验法。

上述讨论的是当方差已知时，一个正态总体均值检验的方法。所采用的检验统计量均是 U—检验统计量 $U = (\overline{X} - \mu_0)\sqrt{n} / \sigma$ 。事实上，采用 U—检验法还可应用于其他的参数估计。

（a）在方差已知的条件下两个正态总体均值差的检验 ：

设二个总体 $X \sim N(\mu_1, \sigma_1^2)$， $Y \sim N(\mu_2, \sigma_2^2)$， $(X_1, \cdots, X_m)$，$(Y_1, \cdots, Y_n)$ 分别是 X 与 Y 的样本。在 σ_1^2， σ_2^2 已知条件下，讨论二正态总体均值差 $\delta = \mu_1 - \mu_2$ 的检验，有基中常见的假设：（1°）H_0： $\delta = 0$， H_1： $\delta \ne 0$； （2°）H_0'： $\delta \ge 0$， H_1'： $\delta < 0$；（3°）H_0''：$\delta \le 0$，H_1''：$\delta > 0$。对这几种假设的检验方法完全类似于上面讨论的方法，所不同的是，此时采用的检验统计量可取为：

$$U = (\overline{X} - \overline{Y} - \delta) \big/ (\sigma_1^2 / m + \sigma_2^2 / n)^{1/2}$$

其中 $\overline{X}, \overline{Y}$ 分别是 X, Y 的样本均值。这里只要注意当 H_0： $\delta = 0$ 时， $U \sim N(0, 1^2)$ 就足以

明白。

（b）对总体非正态分布的大样本情况下均值的检验，亦可采用 U 检验法。例如，总体 $X\sim F(x,\theta)$ 为一般 r.v.，若 $DX=\sigma^2$ 已知，$EX=\theta$ 未知，要检验假设 H_0：$\theta=\theta_0$，H_1：$\theta\neq\theta_0$，在大样本的情况下，可用 $U_n=(\overline{X}-\theta_0)\sqrt{n}/\sigma$ 作为检验统计量。这是因为，在 H_0 成立时，由中心极限定理，U_n 依分布收敛于 $N(0,1^2)$，即 H_0 成立时，U_n 的分布趋向于 U 的分布（即标准正态分布）。故当 n 充分大时,可用 U 作为 U_n 的近似。

注：上例中，若 $DX=\sigma^2$ 未知，可将 U_n 中的σ换成 S，即取 $U_n=(\overline{X}-\theta_0)\sqrt{n}/S$ 作为检验统计量，可以证明，当 n 充分大时，有类似的结论。

（2）当方差 σ^2 未知时，均值的检验 ：

设总体 $X\sim N(\mu,\sigma^2)$，$X_1,\cdots,X_n$ 为其样本，$\overline{X},S^2$ 分别为样本均值与样本方差。若 σ^2 未知，要对如同上面(1)中的三种假设进行检验如下：

（1°）检验 H_0：$\mu=\mu_0$，H_1：$\mu\neq\mu_0$；

由于 σ^2 未知，故而此时的检验统计量选为 $T(n-1)=(\overline{X}-\mu_0)\sqrt{n}/S$，由本章第一节易知，当 H_0：$\mu=\mu_0$ 成立时，$T(n-1)=(\overline{X}-\mu_0)\sqrt{n}/S\sim t(n-1)$，于是，当 $|T(n-1)|$ 取得很小时，应接受 H_0；反之当 $|T(n-1)|$ 取得很大时，应拒绝 H_0。若给定水平α，可查 t 分布表分位数 $t_{\alpha/2}(n-1)$，使得 $P\{|T(n-1)|\leq t_{\alpha/2}(n-1)\}=1-\alpha$，故：

$$\text{接受域}\quad W_0=(\mu_0-t_{\alpha/2}(n-1)S\big/\sqrt{n},\mu_0+t_{\alpha/2}(n-1)S\big/\sqrt{n})$$

$$\text{拒绝域}\quad W_1=(-\infty,\mu_0-t_{\alpha/2}(n-1)S\big/\sqrt{n})\bigcup(\mu_0+t_{\alpha/2}(n-1)S\big/\sqrt{n},+\infty)$$

当$T(n-1)\in W_0$时，接受H_0，否则拒绝之。若检验统计量φ在H_0成立时服从t分布，称为t检验。

（2°）H_0'： $\mu\ge\mu_0$， H_1'： $\mu<\mu_0$；

检验统计量同（1°）取为$T=(\overline{X}-\mu_0)\sqrt{n}/S$。显见，当$T(n-1)$取充分大时，应接受$H_0'$，否则拒绝$H_0'$。给定$\alpha$时，可取：

$$\Phi：W_0'=(\mu_0-t_\alpha(n-1)S\big/\sqrt{n},+\infty)，\ W_1'=(-\infty,\mu_0-t_\alpha(n-1)S\big/\sqrt{n})$$

当$T(n-1)\in W_0'$时，接受H_0'，否则拒绝之。

（3°）H_0''： $\mu\le\mu_0$， H_1''： $\mu>\mu_0$。

请读者给出一个适当的检验法。

不难看到，对于二个正态总体$X\sim N(\mu_1,\sigma_1^2)$， $Y\sim N(\mu_2,\sigma_2^2)$，$X$ 与 Y 的样本为$(X_1,\cdots,X_m)$及$(Y_1,\cdots,Y_n)$。若σ_1^2，σ_2^2未知，但$\sigma_1^2=\sigma_2^2$时，要检验其均值差$\delta=\mu_1-\mu_2$是否为 0，≥ 0，≤ 0三种情况，均可应用t检验法，此时检验统计量可选为：

$$T=\left(\frac{nm}{m+n}\right)^{\frac{1}{2}}(\overline{X}-\overline{Y}-\delta)/S$$

其中$S^2=\dfrac{1}{m+n-2}(\sum\limits_{i=1}^{m}(X_i-\overline{X})^2+\sum\limits_{j=1}^{n}(Y_j-\overline{Y})^2)$。

2. 正态总体方差的检验

1). 一个正态总体方差的检验

设总体 $X \sim N(\mu, \sigma^2)$，其中 μ, σ^2 未知，试由样本 $X_1, \cdots, X_n$ 检验以下问题：

（1°）H_0：$\sigma^2 = \sigma_0^2$，H_1：$\sigma^2 \neq \sigma_0^2$；

（2°）H_0'：$\sigma^2 \geq \sigma_0^2$，H_1'：$\sigma^2 < \sigma_0^2$；

（3°）H_0''：$\sigma^2 \leq \sigma_0^2$，H_1''：$\sigma^2 > \sigma_0^2$。

对（1°）H_0：$\sigma^2 = \sigma_0^2$，已知可取 S^2 作为 σ^2 的估计，因而采取 $K_{n-1}^2 \mathrel{\hat=} (n-1)S^2 / \sigma_0^2$ 作为检验 H_0 的统计量，当 H_0：$\sigma^2 = \sigma_0^2$ 成立时，易知检验统计量 $K_{n-1}^2 \sim \chi^2(n-1)$，对给定的水平α（0<α<1），查卡方分布表定出下分位数 $\chi^2_{1-\frac{\alpha}{2}}(n-1)$ 与上分位数 $\chi^2_{\frac{\alpha}{2}}(n-1)$，可定出

$$\Phi\text{：} W_0 = (\chi^2_{1-\frac{\alpha}{2}}(n-1), \chi^2_{\frac{\alpha}{2}}(n-1))\text{，} W_1 = (0, \chi^2_{1-\frac{\alpha}{2}}(n-1)) \bigcup (\chi^2_{\frac{\alpha}{2}}(n-1), +\infty)$$

当 $K_{n-1}^2 \in W_0$ 时，接受 H_0；否则拒绝 H_0。

以上方法称为卡方检验法。

对（2°）H_0'：$\sigma^2 \geq \sigma_0^2$，Φ：$W_0' = (\chi^2_{1-\alpha}(n-1), +\infty)$，$W_1' = (0, \chi^2_{1-\alpha}(n-1))$。

当 $K_{n-1}^2 \in W_0'$ 时，接受 H_0'；否则拒绝 H_0'。

对（3°）请读者给出一个类似于（2°）的合理的检验法。

问题：某电工厂每天生产一批保险丝，保险丝质量要求之一，是它的熔化时间（单位：小时）的方差 σ^2 越小越好，且规定它不超过 400。对此可提出假设 $H_0'': \sigma^2 \leq 400, H_1'': \sigma^2 > 400$，请读者给该厂提出一个合理的检验方案。

2). 二个正态总体方差比的检验

设 $X \sim N(\mu_1, \sigma_1^2)$，$Y \sim N(\mu_2, \sigma_2^2)$，若 $\mu_1, \sigma_1^2, \mu_2, \sigma_2^2$ 均未知。试由 X 与 Y 的样本 $(X_1, \cdots, X_m)$ 及 $(Y_1, \cdots, Y_n)$ 检验以下问题：

（1°）H_0：$\sigma^2 / \sigma_0^2 = 1$，$H_1$：$\sigma^2 / \sigma_0^2 \neq 1$；

（2°）H_0'：$\sigma^2 / \sigma_0^2 \geq 1$，$H_1'$：$\sigma^2 / \sigma_0^2 < 1$。

对（1°）H_0：$\sigma^2 / \sigma_0^2 = 1$，取检验统计量为 $F = S_1^2 \sigma_2^2 / S_2^2 \sigma_1^2$，则当 $\sigma^2 = \sigma_0^2$ 时，$F \sim F(n-1, m-1)$。给定水平α，取 $F_{1-\frac{\alpha}{2}}(n-1, m-1)$ 与 $F_{\frac{\alpha}{2}}(n-1, m-1)$ 分别为下、上分位数，定出Φ：W_0=($F_{1-\frac{\alpha}{2}}(n-1, m-1)$, $F_{\frac{\alpha}{2}}(n-1, m-1)$)。当 $F \in W_0$ 时，接受 H_0，否则拒绝 H_0。

对（2°）H_0'：$\sigma^2 / \sigma_0^2 \geq 1$，检验统计量同上。Φ：$W_0' = (F_{1-\alpha}(n-1, m-1), +\infty)$，$W_1' = (0, F_{1-\alpha}(n-1, m-1))$。当 $F \in W_0'$ 时，接受 H_0'，否则拒绝 H_0'。

3.指数分布参数的检验

设总体 $X \sim E_X(\lambda)$，试用样本 $X_1, \cdots, X_n$ 检验下列问题：

（1°）H_0：$\lambda = \lambda_0$，H_1：$\lambda \neq \lambda_0$；

（2°）H_0'：$\lambda \geq \lambda_0$，H_1'：$\lambda < \lambda_0$；

（3°）H_0''：$\lambda \leq \lambda_0$，H_1''：$\lambda > \lambda_0$。

（1°）检验 H_0：$\lambda = \lambda_0$，H_1：$\lambda \neq \lambda_0$。注意到在参数估计中，样本均值 $\overline{X}$ 是 $1/\lambda$ 的无偏

估计。且当 $X\sim E_X(\lambda)$时，$2\lambda(\sum_{i=1}^{n}X_i)=2n\lambda\overline{X}\sim\chi^2(2n)$。故取 $K_{2n}=2n\lambda\overline{X}$ 作为检验统计量。给定水平α，查卡方分布表，$\chi^2_{1-\frac{\alpha}{2}}(2n)$ 与 $\chi^2_{\frac{\alpha}{2}}(2n)$，定出检验法Φ：$W_0=(\chi^2_{1-\frac{\alpha}{2}}(2n),\chi^2_{\frac{\alpha}{2}}(2n))$，$W_1=(0,\chi^2_{1-\frac{\alpha}{2}}(2n))\cup(\chi^2_{\frac{\alpha}{2}}(2n),+\infty)$。当 $K_{2n}\in W_0$ 时，接受 H_0，否则拒绝 H_0。

（2°）检验 H_0'：$\lambda\ge\lambda_0$，H_1'：$\lambda<\lambda_0$。取检验统计量同上，Φ：$W_0=(\chi^2_{1-\alpha}(2n),+\infty)$，$W_1=(0,\chi^2_{1-\alpha}(2n))$。当 $K_{2n}\in W_0$ 时，接受 H_0，否则拒绝之。

（3°）检验 H_0''：$\lambda\le\lambda_0$，H_1''：$\lambda>\lambda_0$。请读者自己给出一个检验法。

§4 总体分布的假设检验

本节简要讨论总体分布的假设检验问题。

怎么知道一个总体 X 的分布函数是某个给定的分布函数 $F(x)$呢？这是个在具体应用中经常要遇到的问题，因为许多工程技术、经济管理中的问题所引出的随机变量 X，往往不能根据概率论或有关专业理论推出 X 的（理论）分布函数。只能从得到大样本观测值 $(X_1,\cdots,X_n)=(x_1,\cdots,x_n)$ 的一大堆数据中，用统计推断的方法初步找出统计规律，进而提出总体分布可能是什么样子的猜想或假设。然后，再利用样本观测值对提出的假设进行验证、修改、再验证，以得到较为可靠、满足需要的分布函数。

一般情形大致分为二步。首先对取得的大样本数据作初步整理（包括求 $\bar{x},s^2$，作直方图等），由此推断出总体可能服从的分布函数 $F(x)$（或密度函数 $f(x)$），其次，利用本节介绍的常用非参数检验法检验总体是否就是原假设的分布函数 $F(x)$。

1. 数据初步整理，分布的近似求法

设总体 X 是 *r.v.*，怎样根据样本观测值 $(x_1,\cdots,x_n)$ 近似求出 X 的密度函数（或分布函数）呢？通常可以通过对数据的初步整理与分析，求出 X 的经验分布函数 $\widetilde{F}_n(x)$，作

为对 X 的理论分布 $F(x)$的一个近似（估计）；亦可直接用图解法，例如画直方图（频率图），作为对密度函数的近似（估计）。

数据初步整理与画直方图大致步骤如下：

1. 对样本值 $x_1,\cdots,x_n$ 进行分组：

（1）找出 $x_1,\cdots,x_n$ 的最小值与最大值，记为 $x_{(1)}$ 与 $x_{(2)}$；

（2）取 $a=x_{(1)}$（或略小于 $x_{(1)}$），$b=x_{(2)}$（或略大于 $x_{(2)}$），并等分区间 $[a,b]$ 为 m 个小区间，分点记为 $a=t_0<t_1<\cdots<t_m=b$，其中 $t_{i+1}-t_i=(b-a)/m$。（分组数 m 取多大无确切要求。通常样本容量 n 小时，m 也应小些，但 $m\geq 5$。n 很大时，m 应取大些，例如当 $n\geq 50$ 时，可取 $10\leq m\leq 20$）；

（3）数出样本值落在区间 $(t_{i-1},t_i]$ 上的个数，记为 v_i。$1\leq i\leq m$

2. 计算样本点落在小区间上的频率，即 $f_i=v_i/n$，表样本值落在区间 $(t_{i-1},t_i]$ 上的频率。由大数定律可知，当 n 足够大时

$$p_i=P(t_{i-1}<X\leq t_i)=\int_{t_{i-1}}^{t_i}f(x)dx\approx f(t_i')(t_i-t_{i-1})\quad 其中 t_{i-1}\leq t_i'\leq t_i$$

故 n 充分大时，可取 $f(t_i')\approx p_i/(t_i-t_{i-1})$ 作为 $f(x)$在区间 $(t_{i-1},t_i]$ 上的近似。

3. 画直方图，即在 x—y 坐标平面上，画一排小长方形，对于区间 $(t_{i-1},t_i]$，画以 $(t_{i-1},t_i]$ 为底，以 $y_i=f_i/(t_i-t_{i-1})$ 为高的长方形 $1\leq i\leq m$。如此作出的图叫做直方图。从直方图的形状推测 X 的密度函数 $f(x)$可能是什么样子。

4. 记 $\widetilde{F}_n(x)=\sum_{i=1}^{n}I_{(X_i\leq x)}\big/n,\forall x\in R$，称 $\widetilde{F}_n(x)$ 为 $F(x)$的经验分布，或称为样本的累积频率。由第四、五章可知，$E\widetilde{F}_n(x)$=$F(x)$，且 $\lim\limits_{n\to\infty}\widetilde{F}_n(x)\overset{p}{=}F(x)$（事实上有更强的结果：

$\lim_{n\to\infty}\widetilde{F}_n(x)\overset{a.s.}{=}F(x)$）。因此 $\widetilde{F}_n(x)$ 可作为 $F(x)$的估计量（试验前）和估计值（试验后）。

2. 总体分布的假设检验之一——拟合优度检验

本节讨论如何由总体 X 的样本观测值去检验它是否与某一理论分布相符合，即总体分布函数的假设检验，即 H_0：X 以 $F(x)$为分布函数。

设总体 X 的样本 $X_1,\cdots,X_n$ 的观测值为 $x_1,\cdots,x_n$，按前数据初步整理记号，取 $a=\min_i x_i$，$b=\max_i x_i$。把 $(a,b]$ 用分点 $a=t_0<t_1<\cdots<t_m=b$ 分 $(a,b]$ 为 m 个小区间，v_i 表 $X_1,\cdots,X_n$ 的观察值落在 $(t_{i-1},t_i]$ 上的个数 $1\le i\le m$。则 $f_i=v_i/n$ 表事件 $(t_{i-1}<X\le t_i)$ 在 n 次独立试验中出现的频率。记 $p_i=P(t_{i-1}<X\le t_i)$ 表示事件 $(t_{i-1}<X\le t_i)$ 的概率。若 H_0 成立时，则 $p_i=F(t_i)-F(t_{i-1})$ 就是事件 $(t_{i-1}<X\le t_i)$ 的概率，因此如果 H_0 成立，那么 $(f_i-p_i)^2=(v_i/n-p_i)^2$ 理应比较小。于是

$$K_{m-1}^2=\sum_{i=1}^m (v_i/n-p_i)^2\cdot\frac{n}{p_i}=\sum_{i=1}^m\frac{(v_i-np_i)^2}{np_i}=\frac{1}{n}\sum_{i=1}^m\frac{v_i^2}{p_i}-n \tag{4.1}$$

当 H_0 成立时，也应该取值较小才符合原假设 H_0。换言之，如上式右边每项差异越小，则原假设 H_0 越是可信，也就更加愿意去接受它（式中的因子 $\frac{p_i}{n}$ 在一定程度上起到平衡的作用）。基于上述直观分析，我们可以考虑用上式统计量 K_{m-1}^2 取值的大小作为拒绝或接受 H_0 的推断依据，这个统计量 K_{m-1}^2 称为皮尔逊（*K.Pearson*）的拟合优度检验统计量。这是因为皮尔逊在 1900 年证明了以下重要定理：

定理 4.1 若原假设 H_0（X 以 $F(x)$为分布函数）成立，则当样本容量 $n\to\infty$ 时，K_{m-1}^2 的分布趋向于自由度为$(m-1)$的 χ^2 分布，即 $\chi^2(m-1)$。

这样，当样本容量 n 很大时，可以用 K_{m-1}^2 作为检验 H_0 的检验统计量。下面给出 χ^2 检验法的基本步骤：(1) 原假设 H_0：X 以 $F(x)$ 为分布函数；(2) 检验统计量为 $K_{m-1}^2=\sum_{i=1}^{m}\frac{(v_i-np_i)^2}{np_i}=\frac{1}{n}\sum_{i=1}^{m}\frac{v_i^2}{p_i}-n$；(3) 给定 α，查表 $\chi_\alpha^2(m-1)$，使得 $P(K_{m-1}^2>\chi_\alpha^2(m-1))=\alpha$，定出拒绝域 $W_1=(\chi_\alpha^2(m-1),+\infty)$，接受域 $W_0=(0,\chi_\alpha^2(m-1))$，(4) 计算 K_{m-1}^2 值，当 $K_{m-1}^2\ge\chi_\alpha^2(m-1)$ 时，拒绝 H_0；当 $K_{m-1}^2<\chi_\alpha^2(m-1)$ 时，接受 H_0。

例 4.1 （本例是遗传学上有名的例子）遗传学家孟德尔（*Mendel*）根据对豌豆的大量观测发现豌豆的两对特性——圆与皱，黄与绿所出现的四种组合记录，其频率如下：n=556

组合 I	1 圆黄	2 皱黄	3 圆绿	4 皱绿
v_i	315	101	108	32
p_i	9/16	3/16	3/16	1/16

另一方面，根据他的遗传学理论，孟德尔认为豌豆的上述四种组合的理论推导上的概率为 $p_1=9/16$，$p_2=p_3=3/16$，$p_4=1/16$。现用统计量来检验孟德尔的理论与实测数据的拟合优度。由(4.1)式

$$K_{m-1}^2=\frac{16}{556}(315^2/9+101^2/3+108^2/3+32^2)-556=0.47$$

若给定α=0.05，查表 $\chi_{0.05}^2(3)=7.8$，而 $K_{m-1}^2=0.47<\chi_{0.05}^2(3)=7.8$，故接受 H_0。可知孟德尔的理论分布与实际数据拟合的好。

3. 总体分布的假设检验之二——柯尔莫戈洛夫（*Kolmogorov*）检验

我们在前面介绍了一个总体的经验分布以概率收敛到其理论分布。但事实上，还有更强的结果。只要注意到对固定的 $x\in R$，$F(x)$ 是表示事件 $(X\le x)$ 的概率，而 $\widetilde{F}_n(x)=\sum_{i=1}^{n}I_{(X_i\le x)}/n,\forall x\in R$ 可看作是 n 次独立重复试验中出现事件 $(X\le x)$ 的频率。（注

意 $X_1,\cdots,X_n$ 独立与 X 同分布）。故由贝努里强大数定律，$\forall x\in R$ 固定，有

$$P\{\lim_{n\to\infty}\widetilde{F}_n(x)=F(x)\}=1 \tag{4.2}$$

即 $\widetilde{F}_n(x)$ 以概率 1 收敛到 $F(x)$（$\widetilde{F}_n(x)\xrightarrow{a.s.}F(x)$）。

然而上述结果是针对每一个固定的 $x\in R$ 而言的，因而它的收敛性是局部性态。下面介绍格里文科（*W.Glivenko*）在 1953 年证明的一个重要定理。

定理 4.2 （*W.Glivenko* 定理）设样本 $X_1,\cdots,X_n$ *i.i.d.*~$F(x)$，$\widetilde{F}_n(x)$ 为其经验分布函数，则：

$$P\{\lim_{n\to\infty}\sup_{x\in R}|\widetilde{F}_n(x)-F(x)|=0\}=1 \tag{4.3}$$

即当 $n\to\infty$ 时，经验分布 $\widetilde{F}_n(x)$ 关于 x 均匀地以概率 1 收敛到理论分布 $F(x)$。

以下介绍柯尔莫戈洛夫（*Kolmogorov*）检验

上面介绍利用分布检验总体分布，需将数据进行分组，而分组具有很大的随意性，一方面导致样本信息的无法充分利用。另一方面，所拟合的优度实际上是分组导出的分布，而不是拟合所假设的分布。这里给出柯尔莫戈洛夫检验检验法是利用基于经验分布函数的检验统计量。其基本依据是：由上面的格里文科定理，$n\to\infty$ 时，经验分布 $\widetilde{F}_n(x)$ 以概率 1 一致收敛到总体理论分布 $F(x)$。故可定义 $\widetilde{F}_n(x)$ 到 F 的“距离”为：

$$D_n(\widetilde{F}_n(x),F)=\sup_{x\in R}|\widetilde{F}_n(x)-F(x)| \tag{4.4}$$

现设 X~$F(x)$， $X_1,\cdots,X_n$ 为其样本， $\widetilde{F}_n(x)$ 为其经验分布函数。要检验原假设 H_0：总体 X 的分布 $F(x)=F_0(x)$。其中 $F_0(x)$ 为某一已知分布函数。由(4.3)式知，当 H_0 成立，$n\to\infty$ 时 $D_n(\widetilde{F}_n(x),F)$ 以概率 1 收敛到 0，故而 $D_n(\widetilde{F}_n(x),F)$ 的大小可以显示 $\widetilde{F}_n(x)$ 对总体 $F(x)$ 的拟合程度。D_n 越大，反映拟合越差；反之，D_n 越小，反映拟合越好。因此，可以用

D_n作为检验H_0的检验统计量。为此还需要导出当H_0成立时，D_n的极限分布。因$n\to\infty$时D_n以概率 1 收敛到 0。故考虑极限性态时，需把D_n适当放大$\sqrt{n}$倍，从而考虑$\sqrt{n}D_n$的极限分布。柯氏在 20 世纪 30 年代证明了以下著名定理：

定理 4.3 设总体分布$F(x)$=$F_0(x)$为连续函数，记$D_n=D_n(\widetilde{F}_n,F_0)$，$\forall d>0$，则有：

$$\lim_{n\to\infty}P\{\sqrt{n}D_n<d\}=\sum_{k=-\infty}^{+\infty}(-1)^k\exp(-2k^2d^2) \tag{4.5}$$

证明已超出本书范围，从略。

利用上述结果可以用检验水平为α的检验。取 d，使$\lim\limits_{n\to\infty}P\{\sqrt{n}D_n>d\}=\alpha$，列表如下：

$\sqrt{n}D_n$的临界值（$\lim\limits_{n\to\infty}P\{\sqrt{n}D_n>d\}=\alpha$）

α	0.9	0.75	0.50	0.25	0.10	0.05	0.01
d	0.575	0.678	0.830	1.02	1.23	1.36	1.63

D_n的计算如下：设样本观测值的顺序统计量为$x_{1,n}\le x_{2,n}\le\cdots\le x_{n,n}$，计算

$$d_{k1}=|F_0(x_{k,n})-\frac{k-1}{n}|,d_{k2}=|F_0(x_{k,n})-\frac{k}{n}|,k=1,\cdots,n，\quad D_n=\max\{d_{k1},d_{k2},k=1,\cdots,n\}$$

当$\sqrt{n}D_n$>d时，拒绝H_0；否则接受H_0。

在各种分布函数检验中，K 氏检验的灵敏度高，与分组无关，效果比较好。但此法只能在$F(x)$为连续函数下可用。

练习题

11.1 由经验知某零件重量 $X \sim N(\mu,\sigma^2)$，μ=15, σ^2=0.05。技术革新后，抽了 6 个样品，测得重量为（单位：克）：14.7, 15.1, 14.8, 15.0, 15.2, 14.6，已知方差不变，问平均重量是否仍为 15?（α=0.05)

11.2 糖厂用自动打包机打包，每包标准重量为 100 斤。每天开工后需要检验一次打包机工作是否正常,即检查打包机是否有系统偏差。某日开工后测得几包重量为（单位：斤）如下：

99.3, 98.7, 100.5, 101.2, 98.3, 99.7, 99.5, 102.1, 100.5

问：该日打包机工作是否正常？（α=0.05；已知包重服从正态分布。）

11.3 正常人的脉搏平均为 72 次／分，现某医生测得 10 例慢性四乙基铅中毒患者的脉搏（次／分）如下：

54, 67, 68, 78, 70, 66, 67, 70, 65, 69

问：四乙基铅中毒患者和正常人的脉搏有无显著性差异（已知四乙基铅中毒患者的脉搏服从正态分布。）

11.4 用热敏电阻测温仪间接测量地热勘探井底温度，重复测量 7 次，测得温度（°C）：112.0, 113.4, 111.2, 112.0, 114.5, 112.9, 113.6，而用某精确办法测得温度为 112.6（可看作温度真值），试问用热敏电阻测温仪间接测温有无系统偏差？（α=0.05）

11.5 某种导线，要求其电阻的标准差不得超过 0.005（欧姆）。今在生产的一批导线中抽取样品 9 根，测得 S=0.007(欧姆)，设总体为正态分布。问在水平α=0.05 下能认为这批导线的标准差显著地偏大吗？

11.6 机床厂某日从两台机器所加工的同一零件中，分别抽若干个样测量零件尺寸，得：第一台机器的：6.2, 5.7, 6.5, 6.0, 6.3, 5.8, 5.7, 6.0, 6.0, 5.8, 6.0

第二台机器的：5.6, 5.9, 5.6, 5.7, 5.8, 6.0, 5.5, 5.7, 5.5

问：这两台机器的加工精度是否有显著性差异？（α=0.05）

11.7 某工厂采用新法处理废水，对处理后的水测量所含某种有毒物质的浓度，得到 10 个数据（单位：毫克／升）：

22, 14, 17, 13, 21, 16, 15, 16, 19, 18

而以往用老法处理废水后，该种有毒物质的平均浓度为 19。问：新法是否比老法效果好？（检验水平 α=0.05）

11.8 从正态母体 $N(\mu,1)$中取 100 个样品，计算得 $\bar{x}=10.32$ 。

（1）试检验 H_0：μ=10 是否成立（α=0.01）？

（2）计算上述检验在μ=9.8 时犯第二类错误的概率。

11.9 某电器零件的平均电阻一直保持在 2.64Ω。改变加工工艺后，测得 100 个零件的平均电阻为 2.62Ω，电阻标准差（S）为 0.06Ω，问新工艺对此零件的电阻有无显著影响（α=0.01）？

11.10 某产品的次品率为 0.17。现对此产品进行新工艺试验，从中抽取 400 件检验，发现有次品 56 件。能否认为这项新工艺显著的影响产品的质量（α=0.05）？

11.11 某切割机正常工作时，切割每段金属棒的平均长度为 10.5*cm*。仅在某段时间内随机地抽取 15 段进行测量，其结果如下（*cm*）：10.4，10.6，10.1，10.4，10.5，10.3，10.2，10.9，10.6，10.8，10.5，10.7，10.2，10.7。问该段时间内该机工作是否正常（α=5%）？假定金属棒长度服从正态分布。

11.12 从某种试验物中取出 24 个样品，测量其发热量，计算得 $\bar{x}=11958$，子样标准差 S^*=323，问以 5%的显著水平是否可认为发热量的期望值是 12100（假定发热量服从正态分布）？

11.13 有一种新安眠药，据说在一定剂量下，能比某种旧安眠药平均增加睡眠时间 3 小时。根据资料用旧安眠药睡眠时间平均为 20.8 小时，标准差为 1.6 小时。为了检验这个说法是否正确，收集到一组使用新安眠药的睡眠时间为：

26.7，22.0，24.1，21.0，27.2，25.0，23.4

试问：从这组数据能否说明新安眠药已达到新的疗效（假定睡眠时间服从正态分布，取α=0.05）。

11.14 为了比较两种枪弹的速度（单位：米/秒），在相同条件下进行速度测定。算得子样平均数和子样标准差

枪弹甲 $n_1=110,\quad \bar{x}_1=2805\quad S_1=120.41$

枪弹乙 $n_2=100,\quad \bar{x}_2=2680\quad S_2=105.00$

在显著水平α=0.05 下，这两种枪弹的（平均）速度有无显著差异？

11.15 有甲、乙两台机床加工同样产品，从这两台机床加工的产品中随意地抽取若干件，测得产品直径（单位：*mm*）为

机床甲 20.5，19.8，19.7，20.4，20.1，10.0，19.0，19.9

机床乙 19.7，20.8，20.5，19.8，19.4，20.6，19.2

试比较甲、乙两台机床加工产品直径有无显著差异（α=5%）？假定两台机床加工产品的直径都服从正态分布，且母体方差相等。

11.16 在两个工厂生产的蓄电池中，分别取 10 个测量蓄电池的电容量（单位：安培小时），数据如下：

A 厂 146，141，135，142，140，143，138，137，142，137

B 厂 141，143，139，139，140，141，138，140，142，138

（1）试检验两厂蓄电池的性能有无显著差异（α=5%）？

（2）说明检验要作哪些假定。

11.17 用大子样（n>45）检验在正态母体上作的假设 H_0： $\sigma^2=\sigma_0^2$。试证下面检验方法成立：依据一次抽样后的子样值算得 $\chi^2=(n-1)\dfrac{s^2}{\sigma^2}$ 的数值；若 $\chi^2\ge(n-1)+\sqrt{2(n-1)}u_{\frac{\alpha}{2}}$ ，或 $\chi^2\le(n-1)-\sqrt{2(n-1)}u_{\frac{\alpha}{2}}$ ，则拒绝 H_0 ；若 $(n-1)-\sqrt{2(n-1)}u_{\frac{\alpha}{2}}<\chi^2<(n-1)+\sqrt{2(n-1)}u_{\frac{\alpha}{2}}$，则接受 H_0。

11.18 测的两批电子器材的电阻的子样值为

A 批 x（欧姆）：0.140，0.138，0.143，0.142，0.144，0.137

B 批 y（欧姆）：0.135，0.140，0.142，0.136，0.138，0.140

设这两批器材的电阻分别服从分布 $N(\mu_1,\sigma_1^2)$ 与 $N(\mu_2,\sigma_2^2)$。

（1）检验假设 H_0： $\sigma_1^2=\sigma_2^2$，α=5%;

（2）检验假设 H_0： $\mu_1=\mu_2$，α=5%。

11.19 设总体 X～N (μ,σ^2)，其中 σ^2 已知常数，μ 未知，$X_1,\cdots,X_n$ 为其样本。假设 H_0： $\mu=\mu_0$， H_1： $\mu=\mu_0+\delta$，（$\mu_0,\delta>0$ 为已知常数），给定水平α=0.05。

（1）给出适当的检验;

（2）求检验在 H_1 的功效函数及犯第二类错误的概率;

（3）$\sigma=1,\delta=1$，问 n 应取多大，才能保证第二类错误的概率不超过 0.10。

11.20 设总体 X～N (μ,σ^2)，其中 μ,σ^2 均为未知参数，$X_1,\cdots,X_n$ 是 X 的样本，水平α=0.05。

（1）若 H_0： $\mu=\mu_0$， H_1： $\mu=\mu_0-\delta$，其中 $\delta>0$ 已知，试给出适当的检验;

（2）设 $\delta=1,S=2$，求检验在 H_1 的功效函数及犯第二类错误的概率。

11.21 设总体 X～N (μ,σ^2)，其中 σ^2 已知，μ 为未知参数，$X_1,\cdots,X_n$ 为其样本。给定水平α，任取 $\alpha_i>0,i=1,2,\alpha_1+\alpha_2=\alpha$，记 u_{α_i} 分别为 α_i 的分位数，即

$P\{u > u_{\alpha_i}\} = \alpha_i, i = 1,2$，、易知

$$P\{-u_{\alpha_1} \le (\overline{X} - \mu)\sqrt{n} / \sigma \le u_{\alpha_2}\} = 1 - \alpha \quad (*)$$

试证明在满足(*)式的所有置信度为（1−α）的置信区间$(-u_{\alpha_1}, u_{\alpha_2})$中，当取$\alpha_1 = \alpha_2 = \alpha / 2$时，其置信区间长度最小。

11.22 某车床生产一大批滚珠，随机的抽取 100 个产品，测其直径为（单位：*mm*）：

23.9023, 24.3544, 23.9233, 24.0445, 23.4364,
23.9464, 25.1536, 23.758, 24.0031, 23.8884, 24.0987,
23.0127, 24.5685, 23.5493, 23.757, 22.7177, 23.8448,
22.6404, 24.3034, 24.1123, 23.9506, 23.6562,
24.1476, 23.4952, 22.7377, 24.5511, 23.6283, 24.3536,
24.5179, 23.7127, 23.5198, 24.2968, 24.291, 24.4918,
23.5824, 23.4544, 23.8727, 24.9305, 24.6421, 23.6771,
23.0525, 23.7584, 24.3758, 23.5969, 24.2726, 23.3415,
24.111, 24.0558, 23.4708, 24.5098, 23.7647, 23.3443,
23.7184, 24.075, 23.9409, 24.2786, 23.7967, 24.7233,
23.5952, 23.5276, 24.9612, 24.1335, 23.7452, 22.9702,
22.3605, 24.1839, 24.2938, 23.5494, 23.7727, 23.588,
24.0573, 23.4904, 23.7993, 24.7465, 24.7903, 24.3881,
24.3117, 23.4593, 24.1746, 23.3089, 23.0311, 23.7536,
24.1319, 24.1867, 23.9091, 23.9899, 23.6849, 24.1231,
24.1003, 24.0227, 23.3076, 23.505, 23.5098, 24.5512,
24.2285, 23.5902, 25.1617, 24.5817, 24.0338, 23.3527

（1）用第四节中的数据整理办法对数据进行整理，并画出直方图。

（2）试猜它服从什么分布？并用拟合优度检验法进行检验你的猜想，同时求出该分布的均值，方差。